

**Máster de Formación del Profesorado de Educación Secundaria Obligatoria y
Bachillerato, Formación Profesional y Enseñanza de Idiomas**

Trabajo Fin de Máster

Ámbito Matemáticas

La estadística bidimensional en 4º ESO

Amaia Astarriaga Ros

UNIVERSIDAD PÚBLICA DE NAVARRA

NAFARROAKO UNIBERTSITATE PUBLIKOA

Contenido

Introducción general	5
Parte I: La estadística en el currículo vigente y en los libros de texto	7
1. La estadística en el currículo vigente	11
1.1. Contenidos en Educación Primaria	11
1.2. Contenidos en ESO	14
1.3. Contenidos en Bachillerato	16
2. Criterios de evaluación de la estadística en el currículo vigente	19
2.1. Criterios de evaluación en Educación Primaria	20
2.2. Criterios de evaluación en ESO	21
2.3. Criterios de evaluación en Bachillerato	23
3. Ejercicios, problemas y cuestiones tipo en los libros de texto y su relación con la estadística en el currículo vigente	25
3.1. Ejercicios, problemas y cuestiones tipo en 2º de ESO	25
3.2. Ejercicios, problemas y cuestiones tipo en 3º de ESO	28
3.3. Ejercicios, problemas y cuestiones tipo en 4º de ESO	31
3.4. Ejercicios, problemas y cuestiones tipo en 1º de Bachillerato	33
3.5. Ejercicios, problemas y cuestiones tipo en 2º de Bachillerato	36
4. Resultados	41
4.1. Ausencias y presencias en el currículo y en los libros de texto	41
4.2. Coherencia de los libros de texto en relación con el currículo	47
Parte II: Análisis de un proceso de estudio de la estadística en 4º de ESO	49
5. La estadística en el libro de texto de referencia	53
5.1. Objetos matemáticos involucrados	53
5.2. Análisis global de la unidad didáctica	55
5.3. Otros aspectos relevantes	62
6. Dificultades y errores previsibles en el aprendizaje de la unidad didáctica	67
6.1. Dificultades	67
6.2. Errores y su posible origen	67
7. El proceso de estudio	69
7.1. Distribución del tiempo de la clase	69
7.2. Actividades adicionales planificadas	71

7.3. La tarea: actividad autónoma del alumno prevista.....	71
8. Experimentación	73
8.1. Método.....	73
8.2. Muestra y diseño de la experimentación.....	74
8.3. El cuestionario	74
8.4. Cuestiones y comportamientos esperados	77
8.5. Resultados	78
8.6. Discusión de los resultados	83
9. Síntesis, conclusiones y cuestiones abiertas	85
Síntesis.....	85
Conclusiones.....	85
Cuestiones abiertas	86
Referencias	87
Anexos.....	91
A. Unidad didáctica del libro de texto	93
B. Material didáctico - Apuntes.....	117
C. Material didáctico - Ejercicios	125
D. Prueba	129

Introducción general

Este Trabajo Fin de Máster tiene como objetivo estudiar la enseñanza de la estadística bidimensional en 4º de ESO.

El trabajo se estructura en dos partes. En la primera parte se realiza un estudio longitudinal del currículo y en los libros de texto en Primaria, en ESO y en Bachillerato con relación al tema indicado.

En la segunda parte se propone un proceso de estudio sobre estadística bidimensional, que se ha puesto en marcha en un aula de 4º de ESO en el marco del Practicum II del Máster. Los resultados extraídos de esta experimentación se fundamentan en un cuestionario construido *ad hoc*, teniendo en cuenta asimismo las restricciones institucionales.

El trabajo concluye con una síntesis, unas conclusiones y unas cuestiones abiertas.

Parte I: La estadística en el currículo vigente y en los libros de texto

Esta primera parte del Trabajo Fin de Máster tiene como objetivo analizar cómo se aborda el tratamiento de la Estadística en el currículo y en los libros de texto en Primaria, Secundaria y Bachillerato. Para analizar los libros de texto relativos a cada curso se han revisado los utilizados en el centro de prácticas.

El análisis se divide en cuatro capítulos. En el primer y segundo capítulo se muestran los contenidos y criterios de evaluación del currículo en cada uno de los cursos según la legislación vigente. En el tercer capítulo se exponen ejemplos de actividades tipo propuestas en los libros de texto mencionados para el curso de 4º de ESO, así como en dos cursos anteriores y dos posteriores, clasificadas según su carácter de ejercicio, problema, cuestión o situación.

Las conclusiones que se extraen del análisis comparativo de los contenidos de ambas fuentes (currículo y libro de texto) se exponen en el cuarto capítulo. El objetivo es resaltar las presencias o ausencias de contenido relativas al tema de análisis y valorar la coherencia de los manuales con relación al currículo vigente.

1. La estadística en el currículo vigente

En este capítulo se analiza el contenido de estadística recogido en el currículo oficial para las enseñanzas de Primaria, Secundaria y Bachillerato.

Los contenidos aparecen recogidos en los siguientes decretos del Boletín Oficial del Estado: el Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación Primaria (BOE 293, de 8 de diciembre, 43053-43102); el Real Decreto 1631/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la ESO (BOE 5, de 5 de enero, 677-773) y el Real Decreto 1467/2007, de 2 de noviembre, por el que se establece la estructura del bachillerato y se fijan sus enseñanzas mínimas (BOE 266, de 6 de noviembre, 45381-45477).

1.1. Contenidos en Educación Primaria

Los contenidos de Educación Primaria que aparecen en el BOE relacionados con la estadística corresponden al bloque *Tratamiento de la información, azar y probabilidad*.

Los descriptores se han estructurado en las ramas de la estadística en las que se organiza la educación en primaria y secundaria. En educación primaria tan sólo encontramos contenidos en los descriptores relativos a *C1- Análisis y recogida de información*, *C2- Diagramas y gráficos* y *C3- Medidas de centralización, posición y dispersión*.

Tabla 1

Descriptor	Primaria		
	Primer ciclo	Segundo ciclo	Tercer ciclo
C1- Análisis y recogida de información.	Utilización de técnicas elementales para la recogida y ordenación de datos en contextos familiares y cercanos.	<p>Tablas de datos. Iniciación al uso de estrategias eficaces de recuento de datos.</p> <p>Recogida y registro de datos sobre objetos, fenómenos y situaciones familiares utilizando técnicas elementales de encuesta, observación y medición.</p> <p>Lectura e interpretación de tablas de doble entrada de uso habitual en la vida cotidiana.</p>	Recogida y registro de datos utilizando técnicas elementales de encuesta, observación y medición.
C2- Diagramas y gráficos.	Descripción verbal, obtención de información cualitativa e interpretación de elementos significativos de gráficos sencillos relativos a fenómenos cercanos.	<p>Interpretación y descripción verbal de elementos significativos de gráficos sencillos relativos a fenómenos familiares.</p> <p>Disposición a la elaboración y presentación de gráficos y tablas de forma ordenada y clara.</p>	<p>Distintas formas de representar la información. Tipos de gráficos estadísticos.</p> <p>Valoración de la importancia de analizar críticamente las informaciones que se presentan a través de gráficos estadísticos.</p> <p>Disposición a la elaboración y presentación de gráficos y tablas de forma ordenada y clara.</p> <p>Obtención y utilización de información para la realización de gráficos.</p>
C3- Medidas de centralización, posición y dispersión.	---	---	La media aritmética, la moda y el rango, aplicación a situaciones familiares.
C4- Utilización de las TIC's	---	---	---
C5- Distribuciones bidimensionales.	---	---	---
C6- Inferencia estadística.	---	---	---

En Primaria se introduce la estadística, de manera verbal en el primer ciclo, a través de información cualitativa e interpretación de elementos relativos a fenómenos cercanos al alumnado. Se trata de que desarrollen las herramientas básicas de interpretación de tablas y gráficas que posteriormente van a utilizar en los cursos superiores.

En el segundo ciclo, superada la capacidad de interpretación de tablas y gráficos, se introduce la elaboración de estas por parte de los alumnos. Se trata de gráficas sencillas, con variables cercanas al alumnado que le permitan entender la utilidad de la estadística como herramienta cotidiana.

En el tercer ciclo se introduce el cálculo cuantitativo de parámetros estadísticos elementales de variables cuantitativas y cualitativas sencillas. Además se enseña a valorar críticamente la información recibida a través de gráficos estadísticos.

1.2. Contenidos en ESO

En Educación Secundaria Obligatoria los contenidos del currículo oficial relacionados con estadística corresponden al *Bloque 6. Estadística y probabilidad*, del que se extraen los contenidos propios de estadística organizados según los descriptores utilizados.

Los descriptores utilizados para examinar dichos contenidos son los analizados en primaria a los que se añade *C4- Utilización de las TIC's*.

Tabla 2

Descriptor	Primer Ciclo ESO	
	1º	2º
C1- Análisis y recogida de información.	Diferentes formas de recogida de información. Organización en tablas de datos recogidos en una experiencia. Frecuencias absolutas y relativas.	Diferentes formas de recogida de información. Organización de los datos en tablas. Frecuencias absolutas y relativas, ordinarias y acumuladas.
C2- Diagramas y gráficos.	Diagrama de barras, de líneas y de sectores. Análisis de los aspectos más destacables de los gráficos.	Diagramas estadísticos. Análisis de los aspectos más destacables de los gráficos.
C3- Medidas de centralización, posición y dispersión.	---	Medidas de centralización: media, mediana y moda. Significado, estimación y cálculo. Utilización de las propiedades de la media para resolver problemas. Utilización de la media, la mediana y la moda para realizar comparaciones y valoraciones.
C4- Utilización de las TIC's	---	Utilización de la hoja de cálculo para organizar los datos, realizar cálculos y generar gráficos más adecuados.
C5- Distribuciones bidimensionales.	---	---
C6- Inferencia estadística.	---	---

Tabla 3

Descriptor	Segundo Ciclo ESO		
	3º	4º	
		A	B
C1- Análisis y recogida de información.	<p>Necesidad, conveniencia y representatividad de una muestra. Métodos de selección aleatoria y aplicaciones en situaciones reales.</p> <p>Atributos y variables discretas y continuas.</p>	<p>Identificación de las fases y tareas de un estudio estadístico a partir de situaciones concretas cercanas al alumnado.</p> <p>Análisis elemental de la representatividad de las muestras estadísticas.</p>	<p>Identificación de las fases y tareas de un estudio estadístico.</p> <p>Análisis elemental de la representatividad de las muestras estadísticas.</p>
C2- Diagramas y gráficos.	<p>Agrupación de datos en intervalos. Histogramas y polígonos de frecuencias.</p> <p>Construcción de la gráfica adecuada a la naturaleza de los datos y al objetivo deseado.</p>	<p>Gráficas estadísticas: gráficas múltiples, diagramas de caja.</p>	<p>Gráficas estadísticas: gráficas múltiples, diagramas de caja.</p> <p>Análisis crítico de tablas y gráficas estadísticas en los medios de comunicación. Detección de falacias.</p>
C3- Medidas de centralización, posición y dispersión.	<p>Media, moda, cuartiles y mediana. Significado, cálculo y aplicaciones.</p> <p>Análisis de la dispersión: rango y desviación típica. Interpretación conjunta de la media y la desviación típica.</p> <p>Utilización de las medidas de centralización y dispersión para realizar comparaciones y valoraciones. Actitud crítica ante la información de índole estadística.</p>	<p>Utilización de las medidas de centralización y dispersión para realizar comparaciones y valoraciones.</p>	<p>Representatividad de una distribución por su media y desviación típica o por otras medidas ante la presencia de descentralizaciones, asimetrías y valores atípicos. Valoración de la mejor representatividad en función de la existencia o no de valores atípicos. Utilización de las medidas de centralización y dispersión para realizar comparaciones y valoraciones.</p>
C4- Utilización de las TIC's	<p>Utilización de la calculadora y de la hoja de cálculo para organizar los datos, realizar cálculos y generar las gráficas más adecuadas.</p>	<p>Uso de la hoja de cálculo.</p>	---
C5- Distribuciones bidimensionales.	---	---	---

C6- Inferencia estadística.	---	---	---
-----------------------------	-----	-----	-----

En las tablas se observa la evolución de los contenidos matemáticos relativos a estadística a lo largo del desarrollo de la ESO. En relación con la Educación Primaria vemos que se trata la información estadística desde una perspectiva más crítica. Se estudian los métodos de recogida de información estadística, la representatividad de las muestras escogidas y la necesidad de un análisis crítico de gráficos estadísticos.

En cuanto al desarrollo de contenido matemático, se trabajan las medidas de centralización, posición y dispersión de manera amplia a lo largo de los cursos de la ESO. En el primer ciclo se desarrolla principalmente el cálculo mecánico de estos parámetros y en el segundo ciclo, además, se utilizan para hacer valoraciones y comparaciones.

Por otra parte, en 2º, 3º Y 4º de ESO se introduce la utilización de las TIC's para organizar los datos, realizar cálculos y generar gráficas.

Podemos observar que a lo largo de la ESO hay una evolución de los contenidos estadísticos introducidos, principalmente en la forma de presentarlos. Sin embargo, la introducción de nuevas nociones no tiene apenas variación a lo largo de esta etapa. Vemos que en el segundo ciclo las nociones estadísticas empleadas son las mismas, tanto en 3º, como en 4º A y B.

1.3. Contenidos en Bachillerato

En Bachillerato de Ciencias la estadística apenas tiene cabida, ya que estos dos cursos van enfocados a las Pruebas de Acceso a la Universidad, donde para Matemáticas no hay contenidos estadísticos. En el primer curso de Ciencias, sin embargo, el currículo oficial establece en los contenidos la enseñanza de estadística bidimensional, tratada en un nuevo descriptor, *C5- Distribuciones Bidimensionales*.

Por otra parte, en Ciencias Sociales se da un amplio tratamiento de la estadística, tanto en 1º como en 2º, donde se introduce además el descriptor *C6- Inferencia Estadística*.

Tabla 4

Descriptor	Bachillerato Ciencias	
	1º	2º
C1, C2, C3, C4	---	---
C5- Distribuciones bidimensionales.	Distribuciones bidimensionales. Relaciones entre dos variables estadísticas. Regresión lineal.	---
C6- Inferencia estadística.	---	---

Tabla 5

Descriptor	Bachillerato Ciencias Sociales	
	1º	2º
C1- Análisis y recogida de información.	Estadística descriptiva unidimensional. Tipos de variables. Métodos estadísticos.	Problemas relacionados con la elección de las muestras. Condiciones de representatividad.
C2- Diagramas y gráficos.	Tablas y gráficos.	---
C3- Medidas de centralización, posición y dispersión.	Parámetros estadísticos de localización, de dispersión y de posición.	Parámetros de una población.
C4- Utilización de las TIC's	---	---
C5- Distribuciones bidimensionales.	Distribuciones bidimensionales. Interpretación de fenómenos sociales y económicos en los que intervienen dos variables a partir de la representación gráfica de una nube de puntos. Grado de relación entre dos variables estadísticas. Regresión lineal. Extrapolación de resultados.	Distribuciones de probabilidad de las medias y proporciones muestrales.
C6- Inferencia estadística.	---	Intervalo de confianza para el parámetro p de una distribución binomial y para la media de una distribución normal de desviación típica conocida. Contraste de hipótesis para la proporción de una distribución binomial y para la media o diferencia de medias de distribuciones normales con desviación típica conocida.

Como hemos comentado, en la rama de Ciencias la estadística apenas tiene cabida, sin embargo, en Ciencias Sociales el tratamiento de la estadística en los dos cursos de Bachillerato es importante.

Así como en ESO las nociones estadísticas introducidas apenas tenían variación a lo largo de los cursos, vemos que en Bachillerato, hay un gran salto relativo a la numerosa introducción de nuevo contenido estadístico a lo largo de los dos cursos.

Todos los contenidos desarrollados en la ESO, se vuelven a reforzar y asentar en Bachillerato, donde se estudian de manera más completa y sientan las bases para el posterior estudio de la estadística bidimensional en primero.

En el segundo curso se introducen gran cantidad de nuevos conocimientos al desarrollar íntegramente en este curso la estadística inferencial, con aplicaciones como los intervalos de confianza y contrastes de hipótesis.

2. Criterios de evaluación de la estadística en el currículo vigente

En este capítulo se analizan los criterios de evaluación de estadística recogidos en el currículo oficial para las enseñanzas de Primaria, Secundaria y Bachillerato.

Los criterios de evaluación aparecen recogidos en los siguientes decretos del Boletín Oficial del Estado: el Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación Primaria (BOE 293, de 8 de diciembre, 43053-43102); el Real Decreto 1631/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la ESO (BOE 5, de 5 de enero, 677-773) y el Real Decreto 1467/2007, de 2 de noviembre, por el que se establece la estructura del bachillerato y se fijan sus enseñanzas mínimas (BOE 266, de 6 de noviembre, 45381-45477).

2.1. Criterios de evaluación en Educación Primaria

Tabla 6

Descriptor	Primaria		
	Primer ciclo	Segundo ciclo	Tercer ciclo
E1- Análisis y recogida de información.	---	Recoger datos sobre hechos y objetos de la vida cotidiana utilizando técnicas sencillas de recuento, ordenar estos datos atendiendo a un criterio de clasificación y expresar el resultado en forma de tabla o gráfica.	Realizar, leer e interpretar representaciones gráficas de un conjunto de datos relativos al entorno inmediato.
E2- Diagramas y gráficos.	Realizar interpretaciones elementales de los datos presentados en gráficas de barras. Formular y resolver sencillos problemas en los que intervenga la lectura de gráficos. <i>Con este criterio se trata de valorar la capacidad de interpretar gráficos sencillos de situaciones familiares y verificar la habilidad para reconocer gráficamente informaciones cuantificables.</i>	<i>Este criterio trata de valorar la capacidad para realizar un efectivo recuento de datos y representar el resultado utilizando los gráficos estadísticos más adecuados a la situación. Es asimismo motivo de evaluación la capacidad para describir e interpretar gráficos sencillos relativos a situaciones familiares.</i>	<i>Este criterio trata de comprobar la capacidad de recoger y registrar una información que se puede cuantificar, de utilizar algunos recursos sencillos de representación gráfica: tablas de datos, bloques de barras, diagramas lineales... y de comprender y comunicar la información así expresada.</i>
E3- Medidas de centralización, posición y dispersión.	---	---	---
E4- Utilización de las TIC's	---	---	---
E5- Distribuciones bidimensionales.	---	---	---
E6- Inferencia estadística.	---	---	---

2.2. Criterios de evaluación en ESO

Tabla 7

Descriptor	Primer Ciclo ESO	
	1º	2º
E1- Análisis y recogida de información.	Organizar e interpretar informaciones diversas mediante tablas y gráficas, e identificar relaciones de dependencia en situaciones cotidianas.	Formular las preguntas adecuadas para conocer las características de una población y recoger, organizar y presentar datos relevantes para responderlas, utilizando los métodos estadísticos apropiados y las herramientas informáticas adecuadas.
E2- Diagramas y gráficos.	<i>Este criterio pretende valorar la capacidad de identificar las variables que intervienen en una situación cotidiana, la relación de dependencia entre ellas y visualizarla gráficamente. Se trata de evaluar, además, el uso de las tablas como instrumento para recoger información y transferirla a unos ejes coordenados, así como la capacidad para interpretar de forma cualitativa la información presentada en forma de tablas y gráficas.</i>	<i>Se trata de verificar, en casos sencillos y relacionados con su entorno, la capacidad de desarrollar las distintas fases de un estudio estadístico: formular la pregunta o preguntas que darán lugar al estudio, recoger la información, organizarla en tablas y gráficas,</i>
E3- Medidas de centralización, posición y dispersión.	---	<i>Hallar valores relevantes (media, moda, valores máximo y mínimo, rango) y obtener conclusiones razonables a partir de los datos obtenidos.</i>
E4- Utilización de las TIC's	---	<i>También se pretende valorar la capacidad para utilizar la hoja de cálculo, para organizar y generar las gráficas más adecuadas a la situación estudiada.</i>
E5- Distribuciones bidimensionales.	---	---
E6- Inferencia estadística.	---	---

Tabla 8

Descriptor	Segundo Ciclo ESO		
	3º	4º	
		A	B
E1- Análisis y recogida de información.	Elaborar e interpretar informaciones estadísticas teniendo en cuenta la adecuación de las tablas y gráficas empleadas, y analizar si los parámetros son más o menos significativos. <i>Se trata de valorar la capacidad de organizar, en tablas de frecuencias y gráficas, información de naturaleza estadística, atendiendo a sus aspectos técnicos, funcionales y estéticos (elección de la tabla o gráfica que mejor presenta la información. Asimismo, se valorará la capacidad de interpretar información estadística dada en forma de tablas y gráficas y de obtener conclusiones pertinentes de una población a partir del conocimiento de sus parámetros más representativos.</i>	Elaborar e interpretar tablas y gráficos estadísticos, así como los parámetros estadísticos más usuales correspondientes a distribuciones discretas y continuas, y valorar cualitativamente la representatividad de las muestras utilizadas. <i>Se trata de valorar la capacidad de organizar la información estadística en tablas y gráficas y. En este nivel se pretende, además, que tengan en cuenta la representatividad y la validez del procedimiento de elección de la muestra y analicen la pertinencia de la generalización de las conclusiones del estudio a toda la población.</i>	Elaborar e interpretar tablas y gráficos estadísticos más usuales en las distribuciones unidimensionales y valorar cualitativamente la representatividad de las muestras utilizadas. <i>En este nivel adquiere especial significado el estudio cualitativo de los datos disponibles y las conclusiones que pueden extraerse del uso conjunto de los parámetros estadísticos. Se pretende, además, que se tenga en cuenta la representatividad y la validez del procedimiento de elección de la muestra y la pertinencia de la generalización de las conclusiones del estudio a toda la población.</i>
E2- Diagramas y gráficos.			
E3- Medidas de centralización, posición y dispersión.	<i>Calcular los parámetros centrales (media, mediana, moda) y de dispersión (recorrido y desviación típica) de una distribución.</i>	<i>Calcular los parámetros que resulten más relevantes.</i>	---
E4- Utilización de las TIC's	<i>Utilizar, si es necesario, la calculadora o la hoja de cálculo.</i>	<i>Ayuda de la calculadora o la hoja de cálculo</i>	---
E5- Distribuciones bidimensionales.	---	---	---
E6- Inferencia estadística.	---	---	---

2.3. Criterios de evaluación en Bachillerato

Tabla 9

Descriptor	Bachillerato Ciencias	
	1º	2º
E1- Análisis y recogida de información.	---	Problemas relacionados con la elección de las muestras. Condiciones de representatividad.
E2- Diagramas y gráficos.	---	---
E3- Medidas de centralización, posición y dispersión.	---	Parámetros de una población.
E4- Utilización de las TIC's	---	---
E5- Distribuciones bidimensionales.	Utilizar técnicas estadísticas elementales para tomar decisiones ante situaciones que se ajustan a una distribución de probabilidad binomial o normal. <i>En este criterio se pretende comprobar la capacidad para estimar y asociar los parámetros relacionados con la correlación y la regresión con las situaciones y relaciones que miden.</i>	---
E6- Inferencia estadística.	---	---

Tabla 10

Descriptor	Bachillerato Ciencias Sociales	
	1º	2º
E1- Análisis y recogida de información.	---	<p>Analizar de forma crítica informes estadísticos presentes en los medios de comunicación y otros ámbitos, detectando posibles errores y manipulaciones tanto en la presentación de los datos como de las conclusiones.</p> <p><i>Se valora el nivel de autonomía, rigor y sentido crítico alcanzado al analizar la fiabilidad del tratamiento de la información estadística que hacen los medios de comunicación y los mensajes publicitarios, especialmente a través de</i></p>
E2- Diagramas y gráficos.	---	

		<i>informes relacionados con fenómenos de especial importancia social.</i>
E3- Medidas de centralización, posición y dispersión.	---	---
E4- Utilización de las TIC's	---	---
E5- Distribuciones bidimensionales.	<p>Distinguir si la relación entre los elementos de un conjunto de datos de una distribución bidimensional es de carácter funcional o aleatorio e interpretar la posible relación entre variables utilizando el coeficiente de correlación y la recta de regresión.</p> <p><i>Se pretende comprobar la capacidad de apreciar el grado y tipo de relación existente entre dos variables, a partir de la información gráfica aportada por una nube de puntos; así como la competencia para extraer conclusiones apropiadas, asociando los parámetros relacionados con la correlación y la regresión con las situaciones y relaciones que miden. En este sentido, más importante que su mero cálculo es la interpretación del coeficiente de correlación y la recta de regresión en un contexto determinado.</i></p>	---
E6- Inferencia estadística.	---	<p>Diseñar y desarrollar estudios estadísticos de fenómenos sociales que permitan estimar parámetros con una fiabilidad y exactitud prefijadas, determinar el tipo de distribución e inferir conclusiones acerca del comportamiento de la población estudiada.</p> <p><i>Se pretende comprobar la capacidad para identificar si la población de estudio es normal y medir la competencia para determinar el tipo y tamaño muestral, establecer un intervalo de confianza para μ y p, según que la población sea Normal o Binomial, y determinar si la diferencia de medias o proporciones entre dos poblaciones o respecto de un valor determinado, es significativa.</i></p>

3. Ejercicios, problemas y cuestiones tipo en los libros de texto y su relación con la estadística en el currículo vigente

En este capítulo se hace un análisis de los ejercicios, problemas, cuestiones y situaciones que se plantean en los libros de texto de referencia como reflejo de los contenidos y criterios de evaluación establecidos por el currículo oficial.

Se analizan los libros de texto utilizados en el centro de desarrollo de las prácticas, el colegio Ntra. Sra. del Puy de Estella, relativos a 2º, 3º, 4º de ESO y 1º y 2º de Bachillerato. Los libros de texto de Matemáticas utilizados por el centro son de la Editorial Anaya para todos los cursos, con el criterio de unificar notación y estructura de los contenidos matemáticos, al seguir la misma línea editorial a lo largo de toda la Educación Secundaria.

En el curso de 4º de ESO, en el que se desarrolla la docencia autónoma, no se sigue el libro correspondiente a dicho curso y se proporcionan ejercicios de otros libros de texto, ya que se introduce la estadística bidimensional que en el currículo oficial corresponde a 1º de Bachillerato.

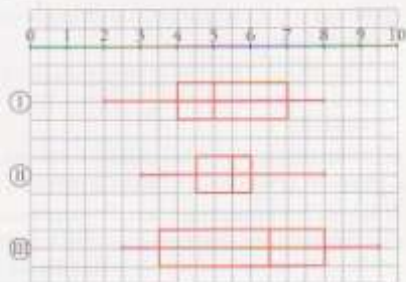
3.1. Ejercicios, problemas y cuestiones tipo en 2º de ESO

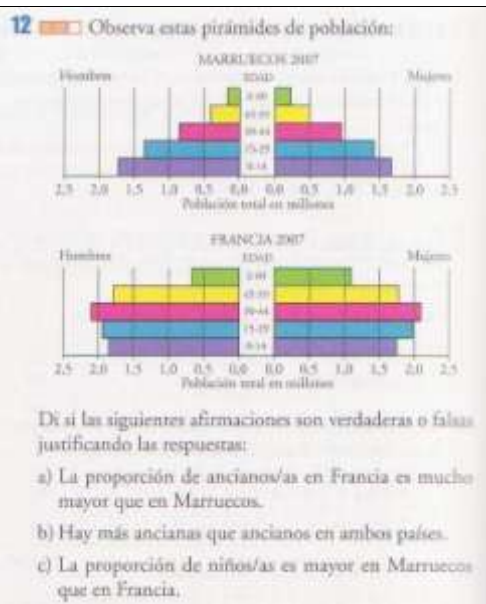
Tabla 11

Libro de texto	Matemáticas 2, Anaya
Tema	12. Estadística
Contenido	<ol style="list-style-type: none"> 1. El proceso que se sigue para realizar estadísticas 2. Tabla de frecuencias 3. Parámetros estadísticos 4. Medidas de posición 5. Gráficas estadísticas 6. Tablas de doble entrada

Tabla 12

Actividad tipo: <input type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input checked="" type="checkbox"/> Cuestión <input type="checkbox"/> Situación
Descripción: Recogida de datos de variables cuantitativas y cualitativas. Se estudian las variables cuantitativas y cualitativas. Los datos se recogen en tablas de frecuencias y se pide razonar si interesa o no calcular la frecuencia relativa o acumulada de los datos.
Ejemplo:
<div> <div>Recogida de datos (Anaya, 2008, 240)</div> <div>Tablas de frecuencias (Anaya, 2008, 243)</div> </div>

<p>Actividades</p> <p>1 Tira tres dados y anota la puntuación intermedia. (Por ejemplo, si sale "1, 2, 6" → anota 2; "3, 5, 5" → anota 5; "1, 1, 4" → anota 1). Realiza la experiencia 20 veces.</p>	<p>Actividades</p> <p>3 La tabla muestra el deporte que prefieren practicar 40 estudiantes.</p> <table border="1"> <thead> <tr> <th>DEPORTE</th><th>FRECUENCIA</th></tr> </thead> <tbody> <tr> <td>Baloncesto</td><td>10</td></tr> <tr> <td>Balonvolea</td><td>1</td></tr> <tr> <td>Fútbol</td><td>20</td></tr> <tr> <td>Tenis</td><td>5</td></tr> <tr> <td>Ajedrez</td><td>4</td></tr> </tbody> </table> <p>Halla las frecuencias relativas de esta distribución y explica por qué carece de sentido hallar las frecuencias acumuladas.</p>	DEPORTE	FRECUENCIA	Baloncesto	10	Balonvolea	1	Fútbol	20	Tenis	5	Ajedrez	4
DEPORTE	FRECUENCIA												
Baloncesto	10												
Balonvolea	1												
Fútbol	20												
Tenis	5												
Ajedrez	4												
<p>Actividad tipo: <input checked="" type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Medidas de centralización, dispersión y posición. Se realiza el cálculo de parámetros estadísticos de manera mecánica.</p> <p>Ejemplo:</p> <p>Parámetros estadísticos (Anaya, 2008, 251)</p> <p>1 Halla la media, la mediana, el recorrido, la desviación media y los cuartiles de las siguientes distribuciones:</p> <p>a) 1, 3, 8, 9, 4, 1, 1, 7, 10, 10</p> <p>b) 1, 3, 5, 4, 2, 8, 9, 6, 10, 6</p>													
<p>Actividad tipo: <input type="checkbox"/> Ejercicio <input checked="" type="checkbox"/> Problema <input type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Comparación de diagramas de caja e interpretación de su significado.</p> <p>Ejemplo: Diagramas de caja (Anaya, 2008, 251)</p> <p>9 a) Compara estas distribuciones de notas obtenidas por tres grupos de alumnos y alumnas indicando cuál es la mediana y los cuartiles Q_1 y Q_3 en cada una de ellas:</p>  <p>b) En la evaluación de estos grupos se hicieron estos comentarios:</p> <ol style="list-style-type: none"> 1. Aprobó el 50% de la clase. 2. Las notas son muy parecidas. 3. La cuarta parte de la clase tiene notas superiores a 7. 4. Es la mejor clase, aunque también es la que tiene mayor dispersión. <p>Indica a qué grupo corresponde cada uno de estos comentarios.</p>													
<p>Actividad tipo: <input type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input checked="" type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Gráficas e interpretación de su significado por medio de cuestiones dirigidas.</p> <p>Ejemplo: Gráficos (Anaya, 2008, 252)</p>													



Actividad tipo: ☐ Ejercicio ☐ Problema ☒ Cuestión ☐ Situación

Descripción: Análisis e interpretación de tablas de doble entrada por medio de cuestiones.

Ejemplo: Tablas de doble entrada (Anaya, 2008, 253)

17 Esta tabla se refiere a los estudiantes de un curso durante el primer trimestre:

	ESTUDIA MENOS DE 2 h DIARIAS	ESTUDIA MÁS DE 2 h DIARIAS	TOTAL
SUSPENDE MÁS DE 2	16	4	
SUSPENDE 0, 1 o 2	2	10	
TOTAL			

Averigua:

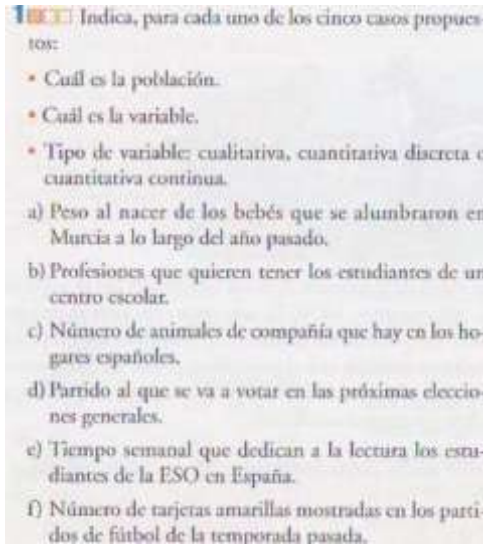
- ¿Cuántos estudiantes hay en total?
- ¿Qué proporción de los estudiantes suspende más de dos asignaturas?
- ¿Qué proporción de los que estudian más de dos horas diarias suspende más de dos asignaturas?
- ¿Qué proporción de los que suspenden más de dos asignaturas estudian más de dos horas diarias?

3.2. Ejercicios, problemas y cuestiones tipo en 3º de ESO

Tabla 13

Libro de texto	Matemáticas 3, Anaya
Tema	12. Estadística
Contenido	<ol style="list-style-type: none"> 1. Población y muestra 2. Variables estadísticas 3. El proceso que se sigue en estadística 4. Confección de una tabla de frecuencias 5. Gráfico adecuado al tipo de información 6. Parámetros estadísticos 7. Cálculo de \bar{x} y σ en tablas de frecuencias 8. Obtención de \bar{x} y σ con calculadora 9. Interpretación conjunta de \bar{x} y σ

Tabla 14

<p>Actividad tipo: <input type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input checked="" type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Cuestiones sobre población y muestra, variables cuantitativas y cualitativas de distintos casos.</p> <p>Ejemplo: Población, muestra, variables (Anaya, 2008, 264)</p> 
<p>Actividad tipo: <input type="checkbox"/> Ejercicio <input checked="" type="checkbox"/> Problema <input type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Problema sobre comparación e interpretación de gráficas.</p> <p>Ejemplo: Interpretación de gráficos (Anaya, 2008, 265)</p>

5 En estos dos diagramas se muestra la composición del organismo en dos edades distintas (aproximadamente):

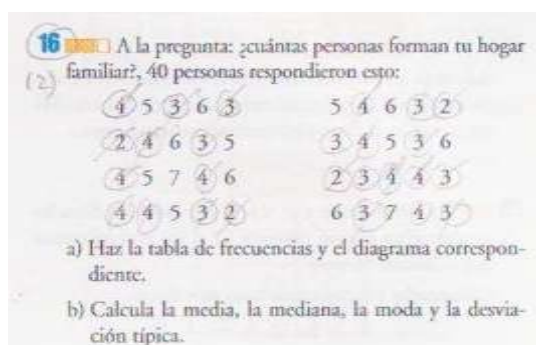


- a) ¿Cómo varía el porcentaje de agua corporal, de masa ósea, de tejido graso y de músculos, órganos...?
- b) Si una persona de 25 años pesa 75 kg, ¿cuál es la cantidad de agua que compone su organismo? ¿Y de tejido graso?
- c) Responde a las preguntas del apartado anterior para el caso de una persona de 75 años con el mismo peso.

Actividad tipo: ☒ Ejercicio ☐ Problema ☐ Cuestión ☐ Situación

Descripción: Elaboración de tablas y gráficos. Se indica el tipo de gráfica que deben realizar los alumnos, lo que no permite que identifiquen por ellos mismos el tipo de gráfico más adecuado a la información dada.

Ejemplo: Elaboración de tablas y gráficos (Anaya, 2008, 267)



Actividad tipo: ☐ Ejercicio ☐ Problema ☒ Cuestión ☐ Situación

Descripción: Cuestión de reflexión sobre la teoría, permite ver si se entienden las nociones de media, desviación típica y coeficiente de variación y la relación entre ellas.

Ejemplo: Comprensión de la teoría (Anaya, 2008, 269)

28 Si dos distribuciones tienen la misma media, y la desviación típica de la primera es mayor que la de la segunda, ¿en cuál de los dos casos es mayor el coeficiente de variación?

Actividad tipo: ☐ Ejercicio ☒ Problema ☐ Cuestión ☐ Situación

Descripción: Problema de proporciones y de medias. Relaciona aspectos del tema con situaciones reales y cercanas al alumnado.

Ejemplo: Problemas (Anaya, 2008, 269)

- 33** Para hallar la nota de una asignatura, el segundo examen vale el doble que el primero, y el tercero, el triple que el primero.
- a) ¿Cuál es la nota final de una alumna que sacó un 5, un 6 y un 4?
- b) ¿Y si esas notas son el 10%, el 40% y el 50% de la nota final, respectivamente?

Otras actividades reseñables

Tabla 15

Actividad tipo: ☐ Ejercicio ☐ Problema ☐ Cuestión ☒ Situación

Descripción: Comparación e interpretación de la misma gráfica en las que se cambia la escala de los ejes.

Relevancia: Se trata de una actividad relevante porque fomenta el sentido crítico de los alumnos. Permite la reflexión sobre la manipulación de las gráficas que habitualmente aparecen en los medios de comunicación condicionadas por el enfoque político que se les quiera dar.

Ejemplo: Población, muestra, variables (Anaya, 2008, 265)

6 Observa esta gráfica en la que aparece la tasa de paro en España desde el año 1995 hasta el año 2000:



Ahora observa esta otra gráfica en la que lo único que ha variado es la escala que se ha tomado en el eje vertical (en la primera iba de 14 a 23, y en la segunda, de 0 a 24):



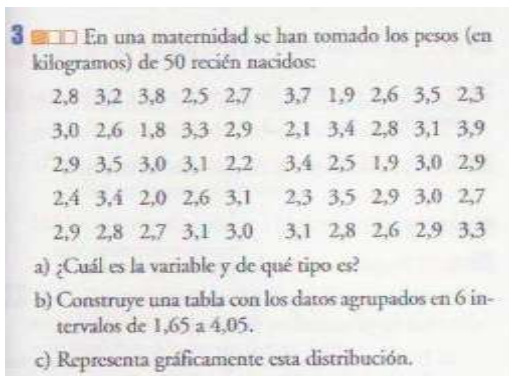
- a) ¿Dan la misma sensación de decrecimiento?
- b) ¿Cuál crees que elegiría el gobierno y cuál la oposición para representar la tasa de paro?





3.3. Ejercicios, problemas y cuestiones tipo en 4º de ESO

Tabla 16

Libro de texto	Matemáticas 4, Anaya
Tema	9. Estadística
Contenido	<ol style="list-style-type: none"> 1. Dos ramas de la estadística 2. Tablas de frecuencias 3. Parámetros estadísticos: \bar{x} y σ 4. Medidas de posición 5. Diagramas de caja 6. Estadística inferencial

Tabla 17

<p>Actividad tipo: <input checked="" type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Agrupar los datos de variables continuas en intervalos y realizar tablas de frecuencias, acompañadas del gráfico más adecuado.</p> <p>Ejemplo: Tablas de frecuencias y gráficos (Anaya, 2008, 201)</p> 
<p>Actividad tipo: <input checked="" type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Ejercicios de cálculo de medias, desviaciones típicas y coeficientes de variación. Se pide comparar datos de dos muestras y razonar los resultados.</p> <p>Ejemplo: Parámetros estadísticos (Anaya, 2008, 201)</p> 
<p>Actividad tipo: <input checked="" type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Ejercicios de medidas de posición, mediana, cuartiles y diagramas de caja en los que se pide razonar e interpretar los resultados.</p> <p>Ejemplo: Medidas de posición (Anaya, 2008, 202)</p>

<p>13  La altura, en centímetros, de un grupo de alumnos y alumnas de una misma clase es:</p> <p>150 169 171 172 172 175 181</p> <p>182 183 177 179 176 184 158</p> <p>Calcula la mediana y los cuartiles y explica el significado de estos parámetros.</p>	
<p>Actividad tipo: <input type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input checked="" type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Realizar tablas de frecuencias de variables continuas, acompañadas del gráfico más adecuado.</p> <p>Ejemplo: Muestreo (Anaya, 2008, 203)</p> <div data-bbox="536 593 1046 969"> <p>25  Para hacer un sondeo electoral en un pueblo de 400 electores, aproximadamente, se va a elegir una muestra de 200 individuos. Di si te parece válido cada uno de los siguientes modos de seleccionarlos y explica por qué.</p> <p>a) Se le pregunta al alcalde, que conoce a todo el pueblo, qué individuos le parecen más representativos.</p> <p>b) Se eligen 200 personas al azar entre las que acuden a la verbena el día del patrón.</p> <p>c) Se seleccionan al azar en la guía telefónica y se les encuesta por teléfono.</p> <p>d) Se acude a las listas electorales y se seleccionan al azar 200 de ellos.</p> </div>	
<p>Actividad tipo: <input type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input type="checkbox"/> Cuestión <input checked="" type="checkbox"/> Situación</p> <p>Descripción: Se tratan temas de inferencia estadística sin recurrir a fórmulas y cálculos, desde la base teórica. Se proponen situaciones en las que los alumnos tienen que razonar la fiabilidad o no de ciertas estimaciones.</p> <p>Ejemplo: Inferencia (Anaya, 2008, 204)</p> <div data-bbox="536 1211 1046 1552"> <p>33  Cada alumno de un grupo cuenta el número de personas y el número de perros que viven en su portal. Suman sus resultados y obtienen una muestra con la que se puede estimar el número de perros que hay en su ciudad.</p> <p>Por ejemplo, supongamos que en su observación obtienen un total de 747 personas y 93 perros. Y saben que en su ciudad viven 75 000 personas.</p> <p>a) ¿Cuántos perros estiman que habrá en la ciudad?</p> <p>b) ¿Cómo es de fiable esta estimación?</p> <p>c) ¿Es aleatoria la muestra que han utilizado?</p> </div>	
<p>Actividad tipo: <input type="checkbox"/> Ejercicio <input type="checkbox"/> Problema <input checked="" type="checkbox"/> Cuestión <input type="checkbox"/> Situación</p> <p>Descripción: Se proponen cuestiones relacionadas con la teoría que permitan a los alumnos razonar sobre las nociones tratadas en el tema.</p> <p>Ejemplo: Reflexión sobre la teoría (Anaya, 2008, 205)</p> <div data-bbox="536 1760 1046 1942"> <p>39  Dos distribuciones estadísticas, A y B, tienen la misma desviación típica.</p> <p>a) Si la media de A es mayor que la de B, ¿cuál tiene mayor coeficiente de variación?</p> <p>b) Si la media de A es el doble que la de B, ¿cómo serán sus coeficientes de variación?</p> </div>	

3.4. Ejercicios, problemas y cuestiones tipo en 1º de Bachillerato

Ciencias sociales

Tabla 18

Libro de texto	Matemáticas aplicadas a las Ciencias Sociales I Bachillerato, Anaya
Tema	8. Estadística
Contenido	<ol style="list-style-type: none"> 1. Estadística. Nociones generales 2. Distribuciones estadísticas 3. Tablas de frecuencias 4. Parámetros estadísticos 5. Parámetros de posición para datos aislados 6. Medidas de posición en distribuciones con datos agrupados en intervalos 7. Interpretación de las medidas de posición. Diagrama de caja

Tabla 19

Actividad tipo: ☒ Ejercicio ☐ Problema ☐ Cuestión ☐ Situación

Descripción: Se proponen ejercicios de resolución sobre los parámetros estadísticos explicados en el tema. Son ejercicios mecánicos en los que tienen que aplicar las fórmulas estudiadas.

Ejemplo: Parámetros estadísticos (Anaya, 2008, 221)

5 En una población de 25 familias se ha observado la variable X = "número de coches que tiene la familia" y se han obtenido los siguientes datos:

0, 1, 2, 3, 1 0, 1, 1, 1, 4 3, 2, 2, 1, 1
 2, 2, 1, 1, 1 2, 1, 3, 2, 1

a) Construye la tabla de frecuencias.
 b) Haz el diagrama de barras.
 c) Calcula la media y la desviación típica.
 d) Halla la mediana y los cuartiles.
 e) Haz un diagrama de caja.

Actividad tipo: ☐ Ejercicio ☐ Problema ☒ Cuestión ☐ Situación

Descripción: Preguntas sobre la teoría para comprobar si se comprenden bien las nociones explicadas en el tema.

Ejemplo: Teoría (Anaya, 2008, 223)

15 En la distribución de las notas de un examen el primer cuartil fue 4. ¿Qué significa esto?

Ciencias sociales / Ciencias

Tabla 20

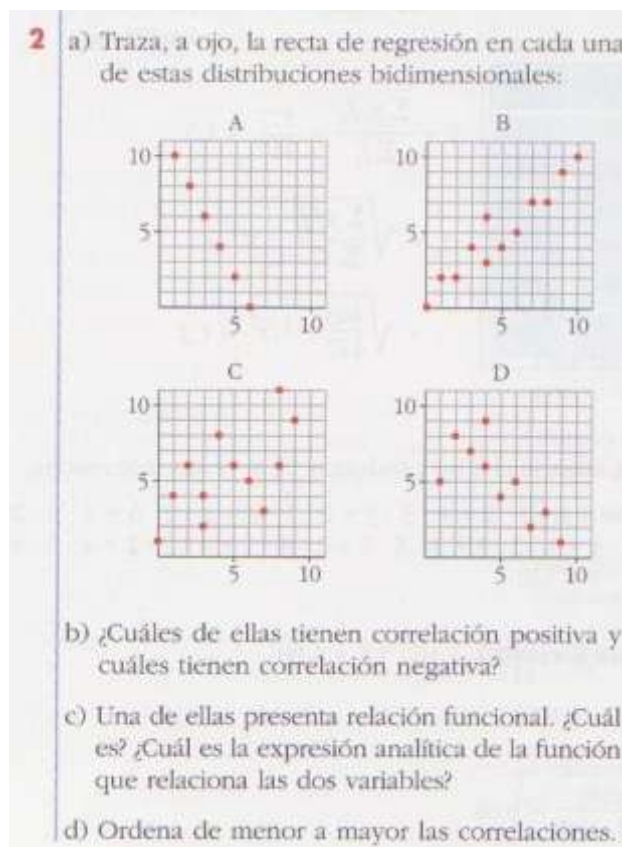
Libro de texto	Matemáticas aplicadas a las Ciencias Sociales I/Matemáticas I Bachillerato, Anaya
Tema	9/13. Distribuciones bidimensionales
Contenido	<ol style="list-style-type: none"> 1. Nubes de puntos. Correlación 2. Medida de la correlación 3. Recta de regresión 4. Hay dos rectas de regresión 5. Tablas de doble entrada

Tabla 21

Actividad tipo: ☒ Ejercicio ☐ Problema ☐ Cuestión ☐ Situación

Descripción: Se proponen cuestiones relacionadas con la teoría en las que hay razonar sobre las nociones tratadas en el tema, sin utilizar fórmulas.

Ejemplo: Recta de regresión (Anaya, 2008, 344)



Actividad tipo: ☐ Ejercicio ☒ Problema ☐ Cuestión ☐ Situación

Descripción: Se proponen problemas de aplicación sobre datos reales, para realizar estimaciones y decir si éstas son fiables o no.

Ejemplo: Estimaciones (Anaya, 2008, 345)

11 La siguiente tabla muestra el número de gérmenes patógenos por centímetro cúbico de un determinado cultivo según el tiempo transcurrido:

Nº DE HORAS	0	1	2	3	4	5
Nº DE GÉRMESES	20	26	33	41	47	53

- a) Calcula la recta de regresión para predecir el número de gérmenes por centímetro cúbico en función del tiempo.
- b) ¿Qué cantidad de gérmenes por centímetro cúbico cabe esperar que haya a las 6 horas? ¿Es buena esta estimación?

Actividad tipo: ☐ Ejercicio ☐ Problema ☒ **Cuestión** ☐ Situación

Descripción: Preguntas sobre la teoría para comprobar si se comprenden bien las nociones explicadas en el tema.

Ejemplo: Teoría (Anaya, 2008, 346)

18 El coeficiente de correlación de una distribución bidimensional es 0,87.

Si los valores de las variables se multiplican por 10, ¿cuál será el coeficiente de correlación de esta nueva distribución?

3.5. Ejercicios, problemas y cuestiones tipo en 2º de Bachillerato

Ciencias sociales

Tabla 22

Libro de texto	Matemáticas aplicadas a las Ciencias Sociales II Bachillerato, Anaya
Tema	11. Las muestras estadísticas
Contenido	<ol style="list-style-type: none"> 1. El papel de las muestras 2. ¿Cómo deben ser las muestras? 3. Tipos de muestreos aleatorios

Tabla 23

Actividad tipo: ☐ Ejercicio ☐ Problema ☒ **Cuestión** ☐ Situación

Descripción: Se proponen cuestiones sobre muestreos en los que tienen que analizar si una muestra es representativa o está sesgada. En otros casos tienen que indicar cuál sería el procedimiento más adecuado para seleccionar una muestra.

Ejemplo: Parámetros estadísticos (Anaya, 2009, 272, 273)

3 Solo uno de los siguientes procedimientos nos permite obtener una muestra representativa. Di cuál es y, en los otros, estudia el sentido del sesgo y su importancia:

a) Para estudiar las frecuencias relativas de las letras, se toman al azar 20 libros de la biblioteca de un centro escolar y se cuenta las veces que aparece cada letra en la página 20 de los libros seleccionados.

b) Para conocer la opinión de sus clientes sobre el servicio ofrecido por unos grandes almacenes de cierta ciudad, se selecciona al azar, entre los que poseen tarjeta de compra, a 100 personas entre las que han gastado menos de 1 000 € el último año, otras 100 entre las que han gastado entre 1 000 € y 5 000 €, y 100 más entre las que han gastado más de 5 000 €.

c) Para calcular el número medio de personas que están adscritas a cada cartilla en un Centro de Salud de la Seguridad Social, los médicos toman nota de todas las cartillas de las personas que acuden a las consultas durante un mes.

8 Queremos seleccionar una muestra de 50 alumnos de 2º de Bachillerato. En cada uno de los siguientes casos debes decidir si el muestreo debe ser aleatorio simple o estratificado por sexos (chicos-chicas) para estudiar las variables indicadas:

a) Estatura.

b) Tiempo que emplean los alumnos en ir de su casa al instituto.

c) Agudeza visual (porcentaje de alumnado con gafas).

d) Incidencia de caries dental.

e) Práctica de fútbol.

f) Lectura de algún periódico.

g) Número de hermanos.

Ciencias sociales

Tabla 24

Libro de texto	Matemáticas aplicadas a las Ciencias Sociales II Bachillerato, Anaya
Tema	12. Inferencia estadística. Estimación de la media
Contenido	<ol style="list-style-type: none"> 1. Distribución normal. Repaso de técnicas básicas 2. Intervalos característicos 3. Distribución de las medias muestrales 4. En qué consiste la estadística inferencial 5. Intervalo de confianza para la media 6. Relación entre el nivel de confianza, error admisible y tamaño de la muestra

Tabla 25

Actividad tipo: ☐ Ejercicio ☐ Problema ☒ **Cuestión** ☐ Situación

Descripción: Se proponen cuestiones sobre intervalos característicos y el teorema central del límite.

Ejemplo: Intervalos característicos (Anaya, 2009, 294)

2 En una distribución normal con media $\mu = 25$ y desviación típica $\sigma = 5,3$, obtén un intervalo centrado en la media, $(\mu - k, \mu + k)$, de forma que el 95% de los individuos estén en ese intervalo.

6 Una variable aleatoria x se distribuye normal $N(120, 30)$. ¿Qué se puede afirmar de la distribución de las medias \bar{x} de las muestras de tamaño n :

a) si $n = 36$
b) si $n = 16$?

Actividad tipo: ☐ Ejercicio ☒ **Problema** ☐ Cuestión ☐ Situación

Descripción: **Problemas sobre estimaciones, se plantean de distintas maneras, realizar una estimación con un nivel de confianza fijado o hallar la probabilidad de que un dato se encuentre en un intervalo prefijado.**

Ejemplo: Intervalos característicos (Anaya, 2009, 295)

12 En una ciudad, la altura de sus habitantes tiene una desviación típica de 8 cm. Si la altura media de dichos habitantes fuera de 175 cm, ¿cuál sería la probabilidad de que la altura media de una muestra de 100 individuos tomada al azar fuera superior a 176 cm?

20 Para estimar la estatura media de los jóvenes entre 15 y 25 años de una localidad, se ha medido a 40 de estos jóvenes, obteniéndose los siguientes resultados:

ESTATURA (cm)	[148, 153)	[153, 158)	[158, 163)
N.º DE JÓVENES	2	4	11

ESTATURA (cm)	[163, 168)	[168, 173)	[173, 178)
N.º DE JÓVENES	14	5	4

Estima, con un nivel de confianza del 99%, el valor de la estatura media de los jóvenes entre 15 y 25 años de dicha localidad.

Ciencias sociales

Tabla 26

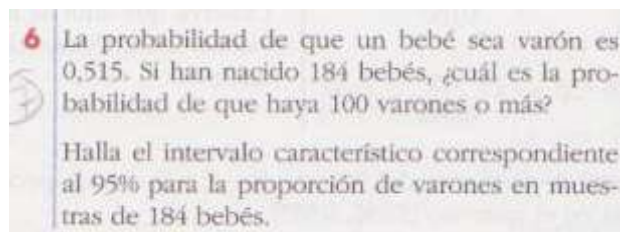
Libro de texto	Matemáticas aplicadas a las Ciencias Sociales II Bachillerato, Anaya
Tema	13. Inferencia estadística: Estimación de una proporción
Contenido	<ol style="list-style-type: none"> 1. Distribución binomial. Repaso de técnicas básicas para el muestreo 2. Distribución de las proporciones muestrales 3. Intervalo de confianza para una proporción o una probabilidad

Tabla 27

Actividad tipo: ☒ Ejercicio ☐ Problema ☐ Cuestión ☐ Situación

Descripción: Ejercicios sobre distribuciones de proporciones muestrales. Se pide hallar la probabilidad de que una proporción esté entre valores dados o el intervalo para un porcentaje prefijado.

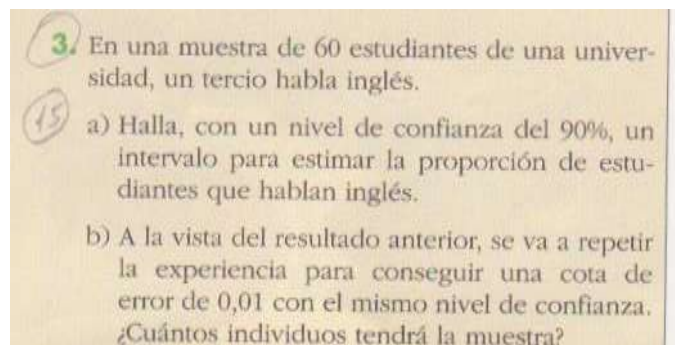
Ejemplo: Distribuciones de proporciones muestrales (Anaya, 2009, 308)



Actividad tipo: ☐ Ejercicio ☒ Problema ☐ Cuestión ☐ Situación

Descripción: Problemas sobre intervalos de confianza para proporciones muestrales. Se pide hallar la intervalos de confianza para un nivel de confianza dado o calcular el tamaño de la muestra necesario para un error máximo fijado.

Ejemplo: Intervalos de confianza (Anaya, 2009, 308)



Ciencias sociales

Tabla 28

Libro de texto	Matemáticas aplicadas a las Ciencias Sociales II Bachillerato, Anaya
Tema	14. Inferencia estadística: Contrastes de hipótesis
Contenido	<ol style="list-style-type: none"> 1. Hipótesis estadísticas 2. Contrastes de hipótesis para la media 3. Contrastes de hipótesis para la proporción 4. Posibles errores en el contraste de hipótesis

Tabla 29

Actividad tipo: ☐ Ejercicio ☒ **Problema** ☐ Cuestión ☐ Situación

Descripción: Se proponen problemas contrastes de hipótesis para la media y para la proporción. Los problemas planteados muestran situaciones reales de contrastes de hipótesis.

Ejemplo: Inferencia estadística (Anaya, 2009, 323)

52 Un fabricante de lámparas eléctricas está ensayando un nuevo método de producción que se considerará aceptable si las lámparas obtenidas por este método dan lugar a una población normal de duración media 2400 horas, con una desviación típica igual a 300.

Se toma una muestra de 100 lámparas producidas por este método, y esta muestra da una duración media de 2320 horas. ¿Se puede aceptar la hipótesis de validez del nuevo proceso de fabricación con un riesgo igual o menor al 5%?

6 Una empresa de productos farmacéuticos afirma en su publicidad que uno de sus medicamentos reduce considerablemente los síntomas de la alergia primaveral en el 90% de la población.

Una asociación de consumidores ha experimentado dicho fármaco en una muestra de 200 socios de la misma, obteniendo el resultado indicado en la publicidad en 170 personas.

Determina si la asociación de consumidores puede considerar que la afirmación de la empresa es estadísticamente correcta al nivel de significación de 0,05.

4. Resultados

En este capítulo se analizan la coherencia de los libros de texto analizados y el currículo exigido por las directrices del Ministerio de Educación en relación a un libro de referencia de estadística descriptiva.

Los libros de texto analizados son los mencionados en el capítulo 3 y referenciados en el anexo, así como los Reales Decretos del BOE.

El libro de matemáticas de referencia que se utiliza es *Estadística*, de Spiegel y Stephens, Editorial Mc Graw Hill, referenciado en el anexo y recomendado por José Antonio Moler, director de este Trabajo.

4.1. Ausencias y presencias en el currículo y en los libros de texto

Al analizar las unidades de estadística de los libros de texto se ve que a lo largo de los cursos se repiten las mismas nociones, pero desde un punto de vista superior, a las que se añaden conceptos y nociones más complejas. El diseño del currículo sigue un aprendizaje en espiral en el que cada año se repasa lo anterior, explicándolo de manera más rigurosa, a la vez que se añade algún nuevo apartado.

Se puede observar que en los cursos de la ESO se introducen progresivamente las siguientes nociones, con el esquema de currículo en espiral. Se abordan los aspectos teóricos de individuo, población y muestras, tipos de variables estadísticas y los pasos para elaborar una estadística. Se calculan los parámetros estadísticos manualmente, a partir de tablas de frecuencias y con la aplicación de las fórmulas correspondientes. Además se trabajan distintos tipos de gráficos estadísticos y la adecuación del uso de cada uno de ellos a las variables estudiadas, así como el tipo de información que aportan.

En los tres cursos analizados, 2º, 3º y 4º de ESO, se estudian estas mismas nociones, añadiendo en cada curso nuevos contenidos. En 3º se añaden en las medidas de dispersión el cálculo de la varianza y la desviación típica. En 4º se introduce la estadística inferencial de manera teórica, se aproxima al alumno a estos nuevos conceptos, cómo tienen que ser las muestras para que las conclusiones sean fiables y que los datos averiguados de la población a partir de una muestra son aproximados, pero que se puede medir el grado de aproximación y el nivel de seguridad.

Podemos observar que la estadística en los cursos de ESO es muy similar e introduce pocos conceptos nuevos a lo largo de los tres cursos analizados. Cambia la manera de exponer la información, y el tipo de ejercicios y problemas que van siendo más complejos en cada curso, pero las nociones teóricas son similares en 2º, 3º y 4º.

Así como en la ESO la estadística estudiada apenas varía a lo largo de los cursos, en Bachillerato se introducen numerosos conceptos nuevos cada curso. En 1º se introducen las distribuciones bidimensionales, en las ramas de Ciencias y de Ciencias Sociales. En esta última, además se vuelve a ver un tema de estadística, que repasa lo dado en ESO de manera más rigurosa. En 2º de Bachillerato sólo aparece la estadística en la rama de Ciencias Sociales, se dedica un tema a

las muestras estadísticas y se introduce y desarrolla la inferencia estadística de manera amplia, a través de tres temas.

Podemos apreciar los contenidos detallados de los temas de estadística en los descriptores de los temas de los libros de texto analizados para Secundaria.

Tabla 30

2º ESO_ Matemáticas 2, Anaya
<p>12. Estadística</p> <ol style="list-style-type: none"> El proceso que se sigue para realizar estadísticas <ul style="list-style-type: none"> Variables estadísticas Tabla de frecuencias <ul style="list-style-type: none"> Frecuencias acumuladas Parámetros estadísticos <ul style="list-style-type: none"> Medidas de centralización Medidas de dispersión Medidas de posición Gráficas estadísticas <ul style="list-style-type: none"> Pictogramas Pirámides de población Climogramas Diagramas de caja Tablas de doble entrada

Tabla 31

3º ESO_ Matemáticas 3, Anaya
<p>12. Estadística</p> <ol style="list-style-type: none"> Población y muestra Variables estadísticas El proceso que se sigue en estadística Confección de una tabla de frecuencias <ul style="list-style-type: none"> Confección de una tabla con datos aislados Confección de una tabla con datos agrupados en intervalos Gráfico adecuado al tipo de información <ul style="list-style-type: none"> Diagrama de barras Histograma de frecuencias Polígono de frecuencias Diagrama de sectores Parámetros estadísticos <ul style="list-style-type: none"> Medidas de centralización Medidas de dispersión Cálculo de \bar{x} y σ en tablas de frecuencias <ul style="list-style-type: none"> Tablas con datos agrupados en intervalos Obtención de \bar{x} y σ con calculadora <ul style="list-style-type: none"> Calculadoras de pantalla sencilla Calculadoras de pantalla descriptiva Interpretación conjunta de \bar{x} y σ <ul style="list-style-type: none"> Coefficiente de variación

Tabla 32

4º ESO_ Matemáticas 4, Anaya	
9. Estadística	
1.	Dos ramas de la estadística
-	Estadística descriptiva
-	Estadística inferencial
2.	Tablas de frecuencias
-	Tabla con datos agrupados
3.	Parámetros estadísticos: \bar{x} y σ
4.	Medidas de posición
-	Frecuencias acumuladas
-	Obtención de percentiles en tablas de frecuencias
5.	Diagramas de caja
6.	Estadística inferencial
-	Por qué se recurre a las muestras
-	Tamaño de la muestra
-	La muestra ha de elegirse al azar
-	Conclusiones que se obtienen de una muestra

Tabla 33

1º Bachillerato. Ciencias _ Matemáticas I Bachillerato Ciencias. Anaya	
13. Distribuciones bidimensionales	
	Reflexiona y resuelve
-	Relación funcional y relación estadística
-	Ejemplo de relación funcional
-	Ejemplo de relación estadística
1.	Nubes de puntos. Correlación
2.	Medida de la correlación
-	Centro de gravedad de una distribución bidimensional
-	Covarianza
-	Correlación
3.	Recta de regresión
-	Método de los mínimos cuadrados
-	La recta de regresión para hacer estimaciones
4.	Hay dos rectas de regresión
-	Posiciones de las dos rectas de regresión
5.	Tablas de doble entrada

Tabla 34

1º Bachillerato. Ciencias sociales_ Matemáticas I Bachillerato Ciencias Sociales. Anaya	
8. Estadística	
1.	Estadística. Nociones generales
-	Dos ramas de la estadística
2.	Distribuciones estadísticas
-	Gráfica adecuada a cada tipo de variable
3.	Tablas de frecuencias
-	Tablas con datos aislados
-	Tablas con datos agrupados en intervalos
4.	Parámetros estadísticos
-	Cálculo e interpretación de \bar{x} y σ
-	Interpretación de los parámetros \bar{x} y σ
5.	Parámetros de posición para datos aislados
-	Mediana y cuartiles
-	Percentiles
-	Frecuencias acumuladas
-	Obtención de percentiles en tablas de frecuencias
6.	Medidas de posición en distribuciones con datos agrupados en intervalos
-	Polígono de frecuencias acumuladas
-	Cálculo de percentiles a partir del polígono de porcentajes acumulados
7.	Interpretación de las medidas de posición. Diagrama de caja
-	Diagrama de caja y bigotes
9. Distribuciones bidimensionales	
1.	Nubes de puntos. Correlación
2.	Medida de la correlación
-	Centro de gravedad de una distribución bidimensional
-	Covarianza
-	Correlación
3.	Recta de regresión
-	Método de los mínimos cuadrados
-	La recta de regresión para hacer estimaciones
4.	Hay dos rectas de regresión
-	Posiciones de las dos rectas de regresión
6.	Tablas de doble entrada

Tabla 35

2º Bachillerato. Ciencias sociales_ Matemáticas II Bachillerato Ciencias Sociales. Anaya	
11. Las muestras estadísticas	<ol style="list-style-type: none"> El papel de las muestras <ul style="list-style-type: none"> Población y muestra ¿Por qué se recurre a las muestras? ¿Cómo deben ser las muestras? <ul style="list-style-type: none"> Muestreo Muestreo aleatorio Tipos de muestreos aleatorios <ul style="list-style-type: none"> Muestreo aleatorio simple Muestreo aleatorio sistemático
12. Inferencia estadística: Estimación de la media	<ol style="list-style-type: none"> Distribución normal. Repaso de técnicas básicas <ul style="list-style-type: none"> Utilización de la tabla de la normal $N(0,1)$ Cálculo de probabilidades en una $N(0,1)$ Cálculo de probabilidades en una $N(\mu, \sigma)$ Intervalos característicos <ul style="list-style-type: none"> Intervalos característicos en distribuciones $N(0,1)$ Intervalos característicos en distribuciones $N(\mu, \sigma)$ Distribución de las medias muestrales <ul style="list-style-type: none"> Teorema Central del Límite Consecuencias del teorema central del límite En qué consiste la estadística inferencial <ul style="list-style-type: none"> Estimación puntual Estimación por intervalos Intervalo de confianza para la media Relación entre el nivel de confianza, error admisible y tamaño de la muestra <ul style="list-style-type: none"> Hallar el tamaño de la muestra dados E y α Hallar el nivel de confianza conociendo E y n
13. Inferencia estadística: Estimación de una proporción	<ol style="list-style-type: none"> Distribución binomial. Repaso de técnicas básicas para el muestreo <ul style="list-style-type: none"> Distribución binomial La distribución binomial se aproxima a la normal Cálculo de probabilidades en una binomial mediante la aproximación a la normal Distribución de las proporciones muestrales Intervalo de confianza para una proporción o una probabilidad
14. Inferencia estadística: Contrastes de hipótesis	<ol style="list-style-type: none"> Hipótesis estadísticas <ul style="list-style-type: none"> Contraste de hipótesis Pasos para efectuar un contraste de hipótesis Contrastes de hipótesis para la media <ul style="list-style-type: none"> Contraste bilateral: $\mu = \mu_0$ Contraste unilateral: $\mu \leq \mu_0$ o $\mu \geq \mu_0$ Contrastes de hipótesis para la proporción <ul style="list-style-type: none"> Contraste bilateral: $p = p_0$ Contraste unilateral: $p \leq p_0$ o $p \geq p_0$ Posibles errores en el contraste de hipótesis <ul style="list-style-type: none"> Probabilidad de cometer error de un tipo u otro

Si comparamos el contenido del currículo y los libros de texto analizados con el libro de referencia detectamos numerosas presencias en las principales nociones que manejan los manuales y alguna ausencia, sobretudo en el nivel de complejidad de las cuestiones tratadas, en la relación entre los diferentes parámetros estadísticos o en el enfoque que se da a las nociones estadísticas.

Atendiendo al capítulo de variables y gráficas, vemos que en los libros de texto se tratan ampliamente, son nociones que se repiten todos los cursos tratadas de manera más compleja conforme se avanza de nivel.

Asimismo, las distribuciones de frecuencias también son correctamente explicadas en los niveles analizados.

En cuanto a ausencias detectadas en las medidas de tendencia central vemos que en los libros de texto únicamente se hace referencia a la media aritmética, no hay mención de la media geométrica ni armónica, ni a la relación existente entre las tres medias. Por otra parte, no se facilitan las fórmulas para el cálculo de la media y la mediana, ya que los ejercicios propuestos para su cálculo tienen muy pocos datos, y su cálculo manual es rápido y sencillo. Tampoco aparece la relación empírica entre la media, la mediana y la moda en curvas de frecuencia sesgadas, ya que entendemos que son cuestiones de mayor complejidad que no tienen cabida en los niveles educativos de secundaria.

Sin embargo, si que se hace mención en los libros de texto a la interpretación de las medidas de tendencia central y de posición.

Si analizamos las medidas de dispersión vemos que las principales nociones están explicadas, como era de esperar, únicamente detectamos ausencias en nociones más complejas como las propiedades de la desviación estándar, relaciones empíricas entre las medidas de dispersión o correcciones para el cálculo de la desviación debido a la agrupación de los datos en clases.

En el tema de estadística bidimensional se hace referencia a la regresión lineal y al método de mínimos cuadrados, pero no a las relaciones no lineales o a problemas en los que intervienen más de dos variables. En cuanto a la correlación se estudian las nociones de correlación lineal y regresión pero no hay referencia en los libros de texto a el error estándar de estimación o a la correlación de series de tiempo y atributos, ya que son cuestiones más complejas que necesitan de conocimientos superiores para su comprensión.

El tema de inferencia estadística es tratado en el libro de texto de manera amplia, ya que corresponde al último curso de bachillerato y el desarrollo de los alumnos permite tratar cuestiones más complejas y de manera más rigurosa. Sin embargo, se detecta la ausencia del estudio de intervalos de confianza para diferencias y sumas, así como para desviaciones estándar.

De la misma manera los contrastes de hipótesis se trabajan de manera rigurosa en el libro de texto aunque también se detecta la ausencia de contrastes para diferencias muestrales.

Una cuestión a destacar en el análisis de los libros de texto en relación con el libro de texto de referencia es la falta de rigor en la notación de los libros de texto. Tanto en los cursos inferiores como en los superiores no hay una correcta notación de las variables poblacionales y muestrales ni de los estimadores. Los parámetros estadísticos utilizados no hacen esta

distinción, lo que puede provocar errores y confusiones en el posterior estudio de la estadística por parte de los alumnos.

4.2. Coherencia de los libros de texto en relación con el currículo

Los contenidos generales de los libros de texto analizados de la editorial Anaya se ajustan al currículo establecido por el Ministerio de Educación detallado en el capítulo 1 del presente trabajo como observamos en los descriptores de los temas.

Por una parte se detecta contenido adicional que los libros de texto adelantan al currículo oficial exigido, y por otra se detecta alguna ausencia principalmente relacionada con el uso de las TIC's.

Por ejemplo en 2º de ESO se tratan las medidas de posición que no son exigidas en el currículo oficial, pero sin embargo, la utilización de la hoja de cálculo no aparece recogida en el libro específicamente para este tema y sí en el currículo.

En 3º de ESO si hay un capítulo del tema dedicado al uso de la calculadora, pero tampoco aparece la referencia a la hoja de cálculo para organizar datos, realizar cálculos y generar gráficas.

En 4º de ESO aparece una introducción a la inferencia estadística, de manera muy descriptiva, sin apenas uso de cálculos, que no recoge el currículo oficial.

Los contenidos de Bachillerato se ajustan al currículo exigido, aunque destaca la ausencia en el libro de texto de intervalos de confianza y contrastes de hipótesis para la diferencia de medias como recoge el currículo.

Parte II: Análisis de un proceso de estudio de la estadística en 4º de ESO

El objetivo de la segunda parte del Trabajo fin de Máster es analizar el proceso de estudio del tema de estadística impartido durante el periodo de prácticas realizado en el Colegio Ntra. Sra. del Puy en una clase de 4º de ESO.

El análisis se divide en cuatro capítulos. En el primer capítulo se analiza cómo trata el libro el tema de Distribuciones bidimensionales. En el segundo capítulo se muestran las dificultades y errores que previsiblemente van a encontrar los alumnos en el proceso de aprendizaje de la materia. En el tercero se presenta la actividad docente que se propone al alumnado así como su distribución. Por último, analizamos la fase de experimentación y las conclusiones que se extraen.

El tema impartido en curso de 4º de ESO es **Distribuciones bidimensionales**. Este tema no se corresponde con lo establecido por el currículo oficial. El currículo establece para este curso en el bloque de estadística los contenidos mostrados en la Tabla 3 de *identificación de las fases y tareas de un estudio estadístico, análisis de la representatividad de las muestras, gráficas estadísticas, medidas de centralización y dispersión para realizar comparaciones y valoraciones*.

Sin embargo, el departamento de matemáticas del centro, atendiendo a una perspectiva global de la enseñanza matemática en secundaria considera más adecuado para una correcta impartición de la totalidad del currículo situar la docencia de la estadística bidimensional en 4º de ESO en vez de en 1º de Bachillerato como se establece en el currículo oficial.

Se realiza esta modificación atendiendo principalmente a dos aspectos. Por una parte el contenido oficial para 4º de ESO es prácticamente el mismo que los alumnos han visto en 3º de ESO (Tabla 3). Por otra parte, el tema de Distribuciones bidimensionales que corresponde oficialmente a 1º de Bachillerato, en la rama de Ciencias no se da, ya que el tiempo es ajustado y es un contenido que no tiene continuidad en 2º de Bachillerato y por lo tanto no entra en Selectividad. Como consecuencia de ello, si los alumnos de la rama de Ciencias no ven estadística bidimensional en 4º de ESO, no la ven posteriormente en Bachillerato.

Por estos motivos desde el departamento de matemáticas se considera adecuado modificar la programación de los contenidos relativos a matemáticas para una correcta impartición de la totalidad del currículo, atendiendo a una mejor estructuración global de los contenidos de estadística en secundaria.

5. La estadística en el libro de texto de referencia

En este capítulo analizamos el tema de Distribuciones bidimensionales impartido en 4º de ESO. El libro de texto de referencia es el de la editorial Anaya de 1º de Bachillerato, ya que este tema corresponde según el currículo oficial a este curso. Sin embargo, debido a que los alumnos no disponen de dicho texto, el tema se imparte por apuntes y ejercicios anexos que se dan a los alumnos.

Para la realización de este capítulo se utiliza como texto de referencia el artículo *Análisis ontosemiótico de una lección sobre la suma y la resta*, de Juan D. Godino, Vincenç Font y Miguel R. Wilhelmi de 2006.

5.1. Objetos matemáticos involucrados

Para analizar los objetos matemáticos involucrados se hace un análisis basado en las tablas del artículo de referencia mencionado (Godino, Font, & Wilhelmi, 2006), pág. 139.

Tabla 36

LENGUAJE
<p>Verbal</p> <ul style="list-style-type: none"> Media, varianza, desviación típica, covarianza, coeficiente de correlación lineal de Pearson, frecuencia. Variable estadística, distribución bidimensional, diagrama de dispersión /nube de puntos, correlación nula/ débil / fuerte / perfecta, recta de regresión, pendiente, estimación fiable/no fiable.
<p>Gráfico</p> <ul style="list-style-type: none"> Nube de puntos Recta de regresión
<p>Simbólico</p> <ul style="list-style-type: none"> $N, \sum, f_i, \bar{x}, \bar{y}, \sigma_x^2, \sigma_y^2, \sigma_x, \sigma_y, \sigma_{xy}, r, >, <, =, m, x_i, y_i, x, y$

Tabla 37

SITUACIONES
<ul style="list-style-type: none"> Problemas contextualizados en los que hay que realizar una estimación a partir de los datos de dos variables e indicar si dicha estimación es fiable o no fiable. Problemas descontextualizados de cálculo de parámetros estadísticos.

Tabla 38

CONCEPTOS
<p>Previos</p> <ul style="list-style-type: none"> • Población, muestra, caracteres estadísticos cuantitativos/cualitativos, variables estadísticas discretas/continuas. • Media, moda, mediana, rango, varianza, desviación típica.
<p>Emergentes</p> <ul style="list-style-type: none"> • Distribución bidimensional, nube de puntos, coeficiente de correlación lineal de Pearson, recta de regresión. • Correlación nula/ débil / fuerte / perfecta. • Estimación fiable/no fiable.

Tabla 39

PROCEDIMIENTOS
<ul style="list-style-type: none"> • Descontextualización del enunciado del problema. • Contextualización de enunciados descontextualizados. • Paso de los datos a una tabla normalizada. • Representación gráfica de los datos mediante la nube de puntos o diagrama de dispersión. • Aplicación de las fórmulas de cálculo de parámetros estadísticos. • Realización de estimaciones. • Razonamiento de la fiabilidad o no fiabilidad de las estimaciones.

Tabla 40

PROPIEDADES
<ul style="list-style-type: none"> • El conjunto de pares de valores (x_i, y_i) se llama distribución bidimensional. • Cada par de valores se representa como las coordenadas de un punto y el conjunto de todas ellas se llama nube de puntos o diagrama de dispersión. • La correlación entre dos variables puede ser fuerte o débil. • La correlación entre dos variables es positiva o negativa. • El coeficiente de correlación es adimensional. • El coeficiente de correlación (r) está comprendido entre -1 y 1. • Si la correlación es perfecta: $r = -1$ o $r = 1$. • Si la correlación es nula: $r = 0$. • El centro de gravedad de una distribución bidimensional es el punto (\bar{x}, \bar{y}), medias de x e y. • La pendiente de la recta de regresión de Y sobre X es el cociente entre la covarianza y la varianza de x. • Si la correlación es fuerte, la estimación es fiable, mientras que si es débil, la estimación no es fiable.

Tabla 41

ARGUMENTOS
<ul style="list-style-type: none"> • Comprobación de las propiedades en casos particulares. • Justificación de las propiedades, utilizando elementos genéricos. • Cálculo analítico y representación gráfica para explicar las propiedades.

5.2. Análisis global de la unidad didáctica

La unidad didáctica a analizar se basa en el tema 13 del libro de 1º de Bachillerato, Distribuciones bidimensionales. Sin embargo, al no disponer los alumnos de 4º de ESO de dicho libro, se realizan unos apuntes basados en dicho tema que se adaptan al curso de docencia. El libro de texto de referencia y los apuntes citados se encuentran en el Anexo de esta memoria.

El tema del libro de texto comienza con una introducción de la estadística en el que se explica la procedencia del término *regresión* acuñado por Galton (1822-1911), acompañada de una imagen como en el resto de temas.

13

DISTRIBUCIONES BIDIMENSIONALES

El término regresión fue acuñado por **sir Francis Galton** (1822-1911), primo del famoso naturalista **Darwin**. Galton estudió la siguiente distribución bidimensional:

Estatura media de un matrimonio - estatura media de sus hijos adultos

Encontró una correlación fuerte: cuanto mayor era la primera, mayor era la segunda. Es decir, cuanto más altos son los padres, más altos tienden a ser los hijos.

Sin embargo, observó que a padres de estatura muy elevada corresponden hijos altos, pero no tanto. Y a padres muy bajos corresponden hijos no tan bajos. Es decir, parece que la estatura de los hijos se aproxima a los valores medios de la población.

Según Galton, la estatura de los hijos *regresa* hacia la media de la población. De ahí el término *regresión*, que, desde entonces, se utiliza para designar una relación estadística cualquiera.



330

Ilustración 1

(Colera, García, Oliveira, & Santanella, 2008, pág. 330)

El tema se organiza de la siguiente manera:

Tabla 42

1º Bachillerato. Ciencias _ Matemáticas I Bachillerato Ciencias. Anaya	
13. Distribuciones bidimensionales	
Reflexiona y resuelve	
<ul style="list-style-type: none"> - Relación funcional y relación estadística - Ejemplo de relación funcional - Ejemplo de relación estadística 	
1.	Nubes de puntos. Correlación
2.	Medida de la correlación
<ul style="list-style-type: none"> - Centro de gravedad de una distribución bidimensional - Covarianza - Correlación 	
3.	Recta de regresión
<ul style="list-style-type: none"> - Método de los mínimos cuadrados - La recta de regresión para hacer estimaciones 	
4.	Hay dos rectas de regresión
<ul style="list-style-type: none"> - Posiciones de las dos rectas de regresión 	
5.	Tablas de doble entrada

Tras la introducción tenemos una página denominada *Reflexiona y resuelve*, donde aparecen cuestiones previas y ejemplos contextualizados de las nociones que posteriormente se ven en el tema. En este caso se aclaran dos nociones que los alumnos pueden no tener claras como son la relación funcional y estadística, se explican mediante ejemplos cotidianos y cercanos al alumnado.

En el primer apartado del tema se explican los conceptos de *distribución bidimensional*, *nube de puntos* y *correlación*. Las explicaciones se hacen primero mediante ejemplos de tablas y gráficas con los que se explican las nuevas nociones.

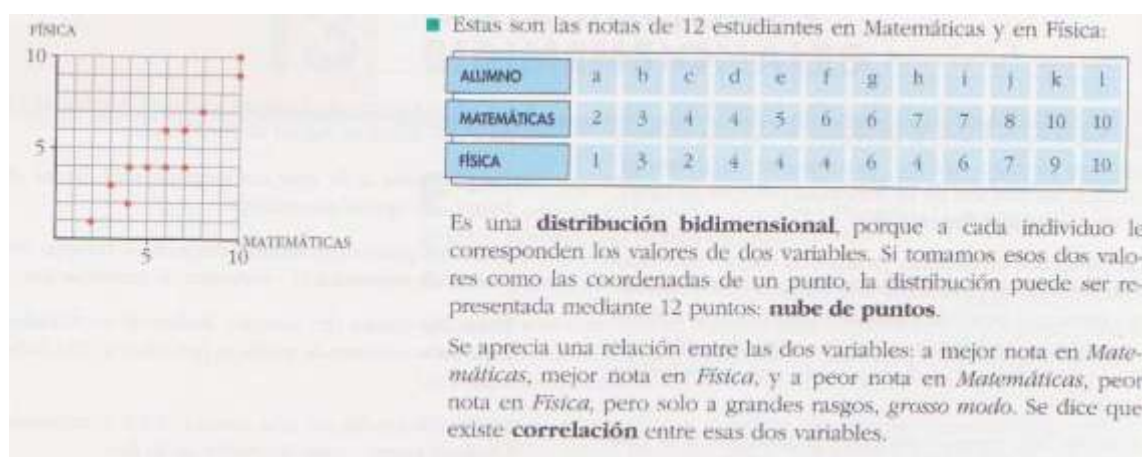


Ilustración 2

(Colera, García, Oliveira, & Santanella, 2008, pág. 332)

Posteriormente se dan las definiciones formales con la notación matemática involucrada.

Al final de este apartado tenemos un ejercicio propuesto para aplicar lo estudiado.

EJERCICIOS PROPUESTOS

1. La tabla de la derecha muestra cómo se ordenan entre sí diez países, A, B, C..., según dos variables, R.P.C. (*renta per cápita*) e I.N. (*índice de natalidad*). Representa los resultados en una nube de puntos, traza la recta de regresión y di cómo te parece la correlación.

PAISES	A	B	C	D	E	F	G	H	I	J
R.P.C.	1	2	3	4	5	6	7	8	9	10
I.N.	10	6	9	5	7	4	1	3	8	2

Ilustración 3

(Colera, García, Oliveira, & Santanella, 2008, pág. 333)

En el segundo apartado se explica la *Medida de la correlación*. Al igual que en el caso anterior, las explicaciones se hacen mediante ejemplos con tablas y gráficas, que permiten una mejor comprensión de los nuevos conceptos involucrados para, posteriormente facilitar la definición formal. También se añaden gráficas explicativas del signo que pueden tomar los parámetros estudiados, en este caso, del signo de la covarianza.

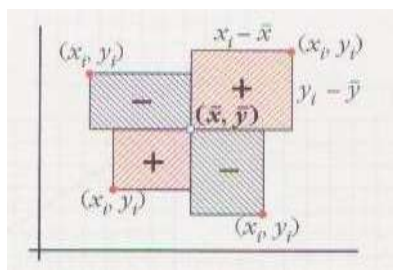


Ilustración 4

(Colera, García, Oliveira, & Santanella, 2008, pág. 334)

Se explican las nociones de *centro de gravedad de una distribución bidimensional*, *covarianza* y *coeficiente de correlación*. Asimismo, se estudia la relación entre la covarianza y la correlación y los valores que puede tomar esta última.

Tras la explicación también tenemos un ejercicio propuesto de aplicación del coeficiente de correlación.

A lo largo de todo el tema tenemos numerosos diagramas que acompañan las explicaciones teóricas, así como recuadros con nociones importantes relativas al tema que debemos recordar y aspectos o puntualizaciones interesantes que los alumnos deben tener en cuenta, se denominan: *ten en cuenta*, *recuerda*, *atención* y *observa*.

TEN EN CUENTA

El centro de gravedad (\bar{x}, \bar{y}) no es, en general, un punto de la distribución.

Ilustración 5

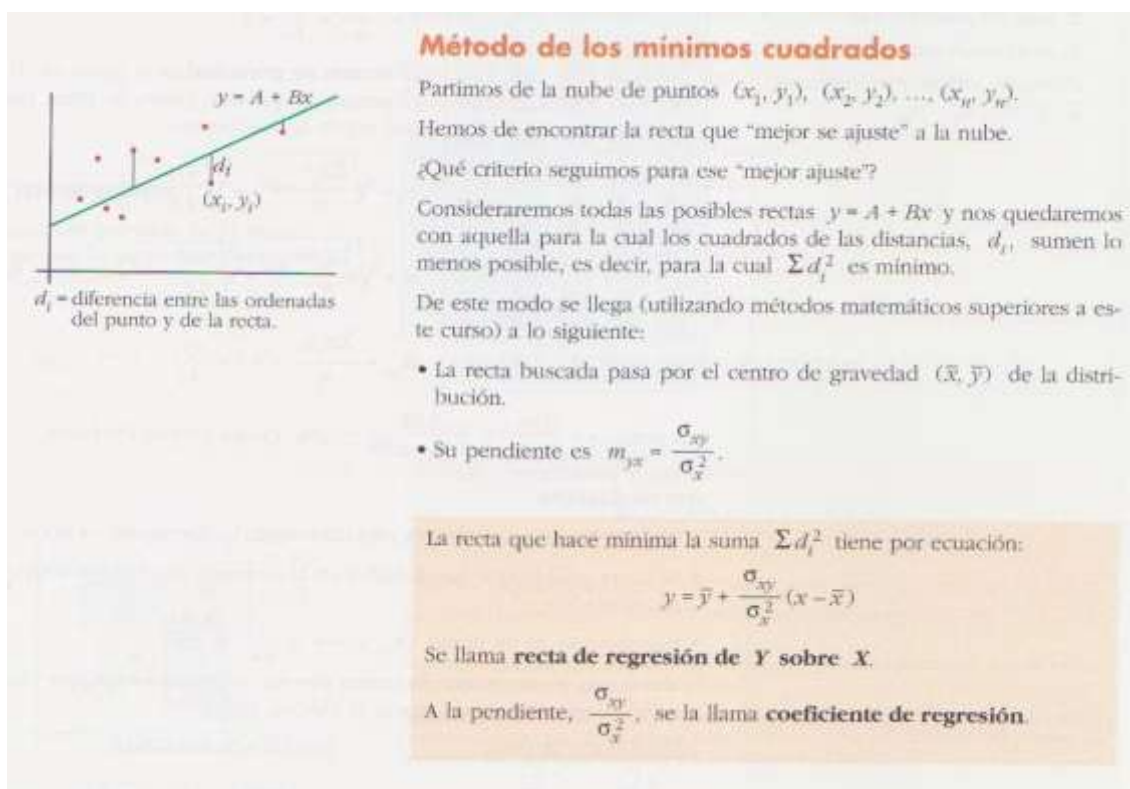
OBSERVA

La recta que nosotros asignamos “a ojo” a una nube de puntos es, aproximadamente, la bisectriz de las dos rectas de regresión.

Ilustración 6

(Colera, García, Oliveira, & Santanella, 2008, pág. 334)

El tercer apartado se dedica al estudio de la recta de regresión. Se menciona que el ajuste de esta recta se realiza mediante el método de mínimos cuadrados, aunque no se profundiza en su explicación, sino que se da directamente el centro de gravedad de la recta y la pendiente, diciendo que estos se calculan “*por métodos matemáticos superiores*”. Tras la definición de la recta se explica su aplicación para hacer estimaciones.

**Ilustración 7**

(Colera, García, Oliveira, & Santanella, 2008, pág. 336)

En este apartado tenemos dos ejercicios resueltos que ayudan a los alumnos a la comprensión de la teoría explicada y a la posterior realización de ejercicios de manera autónoma.

EJERCICIOS RESUELTOS

1. En la distribución *Notas de Matemáticas – Notas de Física*, cuyos parámetros hemos calculado en la página anterior, obtener la recta de regresión de *Y* sobre *X*.

- Pendiente $\rightarrow m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{5,92}{6} = 0,986$
- Ecuación $\rightarrow y = 5 + 0,986(x - 6) \rightarrow y = -0,917 + 0,986x$

CON CALCULADORA

Después de introducir los datos como se ha hecho en la página anterior, se obtiene:

- Pendiente de la recta: $\rightarrow B = 0,986$
- Ordenada en el origen: $\rightarrow A = -0,917$

Ilustración 8

(Colera, García, Oliveira, & Santanella, 2008, pág. 336)

Las dos rectas de regresión posibles (en función de 'x' o de 'y') se explican en el cuarto apartado del tema. Se explican las posiciones relativas que pueden tomar ambas rectas y su significado relacionado con el coeficiente de correlación. Aunque el tema se estructure en apartados diferentes, hay una constante referencia al resto de nociones explicadas en el tema, dotando a éste de la continuidad y conexión apropiadas.

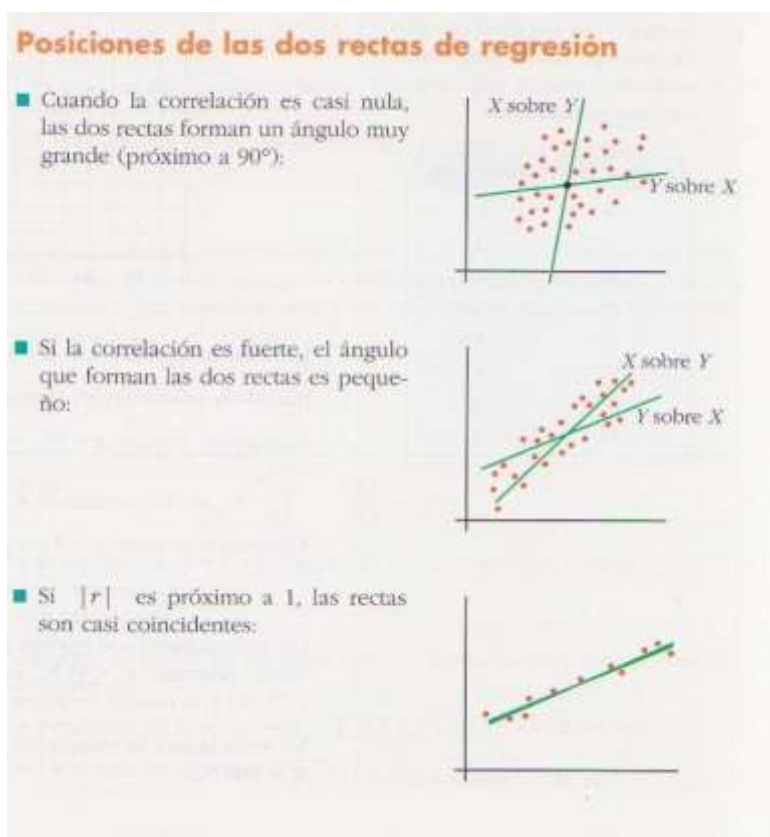


Ilustración 9

(Colera, García, Oliveira, & Santanella, 2008, pág. 338)

Por último, hay un apartado dedicado a las tablas de doble entrada, donde se explican los beneficios de utilizar este tipo de tablas para organizar la información en caso de tener un gran número de datos. Se acompaña de gráficas y barras realizadas a partir de tablas de doble entrada.

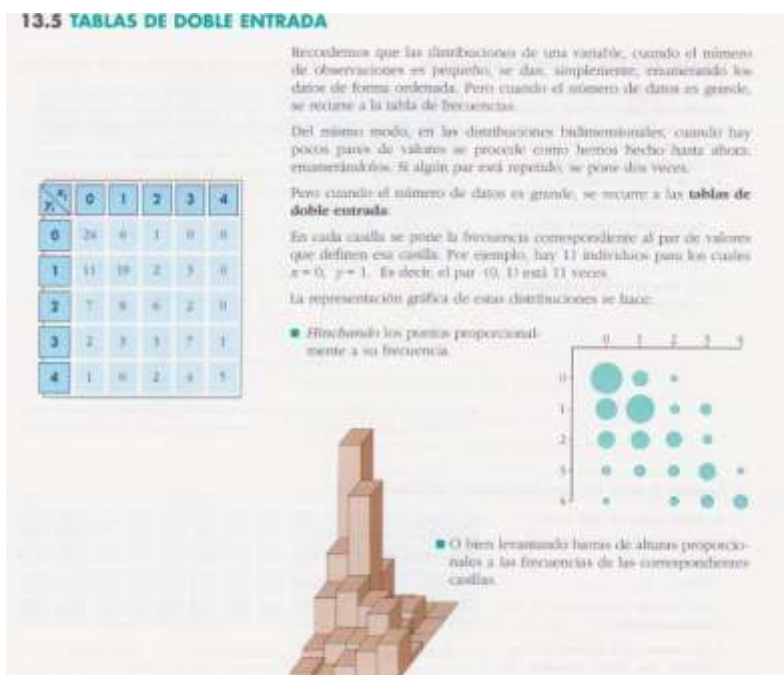


Ilustración 10



(Colera, García, Oliveira, & Santanella, 2008, pág. 339)



Tras la exposición teórica del tema tenemos una serie de ejercicios resueltos con un ejemplo tipo de cada aspecto tratado en el tema y posteriormente numerosos ejercicios y problemas de distinto tipo y de varios niveles de dificultad. El apartado de ejercicios acaba con una autoevaluación que evalúa los aspectos más importantes tratados en el tema y que dispone de las soluciones en el anexo del libro.

- Ejercicios y problemas resueltos
 - Relación funcional y relación estadística
 - Análisis gráfico de una distribución bidimensional
 - Cálculo del coeficiente de correlación
 - Recta de regresión
 - Dos rectas de regresión. Estimación
 - Tabla de doble entrada
- Ejercicios y problemas propuestos
 - Para practicar
 - Sin fórmulas
 - Con fórmulas
 - Para resolver
 - Cuestiones teóricas
 - Para profundizar
- Autoevaluación

AUTOEVALUACIÓN

1. Observa estas distribuciones bidimensionales:

a)  b) 

c)  d) 

Asigna razonadamente uno de los siguientes coeficientes de correlación a cada gráfica:

0,2 -0,9 -0,7 0,6

2. Representa esta distribución bidimensional:

x	1	2	2	3	4	6	7	8	8	9
y	2	4	3	4	6	5	8	9	10	9

a) Calcula los parámetros \bar{x} , \bar{y} , σ_x , σ_y , σ_{xy} .

b) Halla el coeficiente de correlación.

c) Halla la recta de regresión de Y sobre X.

d) Estima el valor de y para $x = 5$ y para $x = 10$. ¿Son "buenas" estas estimaciones?

3. La recta de regresión de Y sobre X de una cierta distribución bidimensional es $y = 1,6x - 3$. Sabemos que $\bar{x} = 10$ y $r = 0,8$.

a) Calcula \bar{y} .

b) Estima el valor de y para $x = 12$ y para $x = 50$. ¿Qué estimación te parece más fiable?

c) Halla la recta de regresión de X sobre Y.

4. El consumo de energía per cápita y en miles de kWh y la renta per cápita x en miles de euros de seis países son:

	A	B	C	D	E	F
x	11,1	8,5	11,3	4,5	9,9	6,5
y	5,7	5,0	5,1	2,7	4,6	3,1

a) Calcula la recta de regresión de Y sobre X.

b) Halla el coeficiente de correlación entre el consumo y la renta.

c) ¿Qué predicción podemos hacer sobre el consumo de energía per cápita de un país cuya renta per cápita es de 4,4 miles de euros?

5. En tu CD puedes encontrar las resoluciones de todos estos ejercicios.

Ilustración 11

(Colera, García, Oliveira, & Santanella, 2008, pág. 347)

5.3. Otros aspectos relevantes

El tema de Distribuciones bidimensionales adaptado a 4º de ESO sufre algunas modificaciones, como se ha comentado en la introducción del capítulo, principalmente por el menor nivel formativo de los alumnos a los que se dirige, pero también debido a la restricción de tiempo (5 sesiones).

Una parte del tema contiene apartados del currículo relativos a 4º de ESO, como el cálculo de parámetros estadísticos y el uso de las medidas de centralización y dispersión para realizar comparaciones y valoraciones.

Por otra parte, se introducen las nociones fundamentales de las distribuciones bidimensionales, pero se simplifican algunos términos y procedimientos. En muchos casos, se procede al cálculo mecánico de los ejercicios, para que los alumnos se familiaricen con las nuevas nociones y expresiones, pero sin hacer mucho hincapié en la fundamentación teórica, ya que en este curso la diversidad del alumnado no permite todavía ahondar en cuestiones teóricas complejas o demasiado abstractas. Una de las simplificaciones es el cálculo de la recta de regresión que se hace únicamente de Y sobre X. Otro aspecto que no se menciona en 4º aunque venga en el tema de 1º de Bachillerato son las tablas de doble entrada, debido a que no se proponen posteriormente ejercicios de este tipo.

Una restricción importante a la hora de planificar la actividad docente ha sido la no utilización del aula de informática debido a las pocas sesiones correspondientes a este tema. Hubiese sido interesante utilizar los ordenadores para poder realizar ejercicios más complejos con el uso de hojas de cálculo, realizar gráficos adecuados a las variables estudiadas, realizar diagramas de dispersión e incluso comprobar el coeficiente de correlación. Ya que en este tema la cantidad de cálculos mecánicos que tienen que hacer los alumnos para realizar estimaciones es alta, por lo que los datos de los ejercicios tienen que ser pocos y sencillos, lo que limita mucho a la hora de realizar ejercicios.

Por otra parte, de esta manera, se fomenta el uso de la calculadora y el cálculo y resolución de ecuaciones, que aunque no corresponde a este tema es un conocimiento transversal que tienen que desarrollar los alumnos. Dada esta circunstancia, se aprovecha el tema para valorar otros aspectos, haciendo hincapié en la importancia del redondeo y la precisión. Se valora en la realización de los ejercicios la precisión de los resultados. Se establece como convenio tomar dos decimales para expresar el resultado de los parámetros, y tres decimales para el coeficiente de correlación, correctamente redondeados. También se insiste en la necesidad de realizar los cálculos con la calculadora, sin meter datos redondeados, desde el principio hasta el final, para que el resultado sea más preciso.

El tema se da en forma de apuntes que se encuentran en el Anexo B y se estructura de la siguiente manera:

- Estadística bidimensional
 - Distribución bidimensional
 - Nube de puntos
 - Correlación
 - Parámetros estadísticos
 - Media aritmética
 - Varianza
 - Desviación típica
 - Covarianza
 - Coeficiente de correlación lineal
 - Recta de regresión de Y sobre X .
 - Aplicación
- Ejemplos y ejercicios

Primero se introduce el tema, diciendo qué es la estadística y para qué sirve, con algún ejemplo verbal de uso cotidiano de la estadística y de la estadística bidimensional en concreto.

Posteriormente se explica qué es una *distribución bidimensional* y se acompaña de un ejemplo por medio de una tabla con dos variables.

1. DISTRIBUCIÓN BIDIMENSIONAL

Se llama **distribución bidimensional** al conjunto de pares de valores $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, donde x e y son dos variables conocidas de una población o muestra de n individuos.

La población es el conjunto de los elementos que queremos estudiar, la muestra es un subconjunto de la población de la que hacemos el estudio. Por ejemplo, población: los habitantes de Estella entre 12 y 16 años, muestra: se seleccionan 100 de estos, de los que se extrae una conclusión.

Se utiliza la estadística bidimensional para ver la relación que hay entre dos variables, por ejemplo, ingresos y gastos de cada familia de una empresa, superficie de un conjunto de países y nº de habitantes, latitudes y temperaturas medias de un grupo de países, notas en dos asignaturas diferentes, etc.

Ejemplo: Las notas de cuatro alumnos en Matemáticas y en Historia son las siguientes:

Matemáticas (x_i)	1	3	6	9
Historia (y_i)	7	5	4	3

Ilustración 12

Tras esto, se define *nube de puntos*, que se representa gráficamente según el ejemplo anterior.

Por medio del ejemplo de la nube de puntos se explica el grado de *correlación* entre dos variables, que posteriormente se generaliza y la correspondencia gráfica de la correlación fuerte, débil o nula. También se muestra mediante el gráfico el significado de la correlación positiva o negativa. En este apartado se hace hincapié en que la relación entre las dos variables no es funcional.

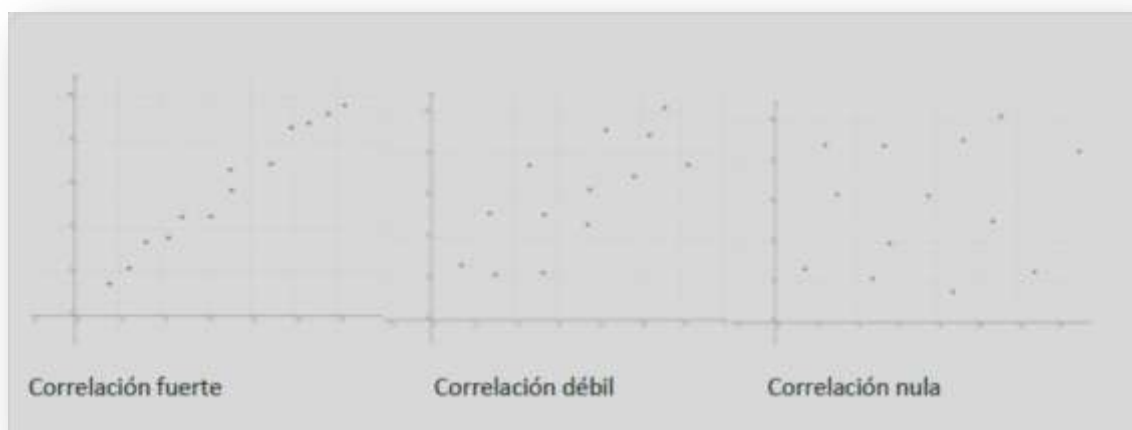


Ilustración 13

Después se recuerda el cálculo de los *parámetros estadísticos* y las fórmulas que se utilizan para ello, con la novedad de que ahora se tienen dos variables X e Y . Se calcula la media aritmética, la varianza y desviación típica y por último la covarianza de una variable bidimensional. El único parámetro que es nuevo en 4º es la covarianza, que relaciona las dos variables X e Y , se indica que esta puede ser positiva o negativa, a diferencia de los parámetros anteriores, que siempre toman valores positivos.

4. PARÁMETROS ESTADÍSTICOS

N = número de datos $N = \sum f_i$ $f_i = 1$

- **Media aritmética**

Se llama media aritmética de una variable estadística a la suma de todos los valores de dicha variable dividida por el número de valores.

$$\bar{x} = \frac{\sum f_i x_i}{N} \quad \bar{y} = \frac{\sum f_i y_i}{N}$$

- **Varianza (var)**

La varianza de una variable es la media aritmética de los cuadrados de las desviaciones respecto a la media.

$$\sigma_x^2 = \frac{\sum f_i x_i^2}{N} - \bar{x}^2 \quad \sigma_y^2 = \frac{\sum f_i y_i^2}{N} - \bar{y}^2$$

- **Desviación típica**

Se llama desviación típica de una variable a la raíz cuadrada positiva de la varianza.

$$\sigma = +\sqrt{\text{var}}$$

$$\sigma_x = +\sqrt{\frac{\sum f_i x_i^2}{N} - \bar{x}^2} \quad \sigma_y = +\sqrt{\frac{\sum f_i y_i^2}{N} - \bar{y}^2}$$

- **Covarianza**

Se llama covarianza de una variable bidimensional (X, Y) a la media aritmética de los productos de las desviaciones de cada una de las variables respecto a sus medias respectivas.

$$\sigma_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y}$$

(La covarianza puede ser positiva o negativa).

Ilustración 14

Hay que tener en cuenta que para simplificar los cálculos (por encontrarnos en 4º curso) y para que las tablas sean más sencillas se toma como frecuencia siempre 1 ($f_i = 1$), y en caso de que aparezca un dato varias veces, se repite poniendo frecuencia uno todas ellas.

A partir del cálculo de la covarianza y de las desviaciones típicas de x y de y, se define otra noción nueva para los alumnos, el *coeficiente de correlación*. Se explica el rango de valores entre los que se mueve y el significado en función del valor que toma.

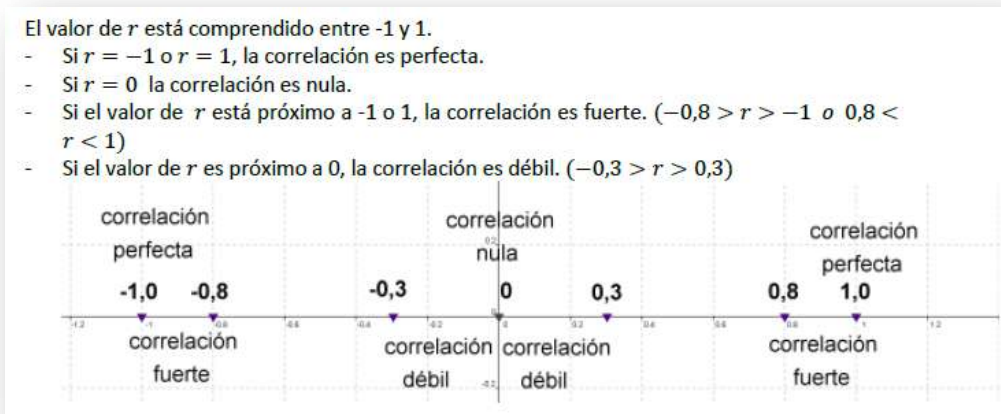


Ilustración 15

Posteriormente se muestra el cálculo de la *recta de regresión de Y sobre X* y su utilidad para hacer estimaciones.

Por último se explica la necesidad de indicar la fiabilidad o no de las estimaciones realizadas según el valor del coeficiente de correlación calculado anteriormente.

El apartado de ejercicios está compuesto por problemas de aplicación. En los ejercicios se pretende que los alumnos sean capaces de realizar estimaciones e indicar la fiabilidad de las mismas. Para ello, primero se procede con ejercicios guiados en los que se indican los pasos a seguir: cálculo de parámetros, cálculo del coeficiente de correlación, recta de regresión, estimación y fiabilidad de la misma. Posteriormente se presentan ejercicios en los que se pide una estimación y los alumnos tienen que deducir el procedimiento para realizarla.

Otras variables que se tienen en cuenta a la hora de elegir los problemas son el sentido positivo o negativo de la correlación y el valor del coeficiente de correlación, siendo esta fuerte o débil. Para que a los alumnos se les haga más fácil predecir la fiabilidad de la estimación, ésta se da en función de los valores que toma el coeficiente de correlación según la recta de la Ilustración 15.

6. Dificultades y errores previsibles en el aprendizaje de la unidad didáctica

En este capítulo se indican las dificultades y los errores que se prevén que cometa el alumno en el estudio del tema de Distribuciones bidimensionales en 4º de ESO.

6.1. Dificultades

Una de las primeras dificultades con las que se encuentran los alumnos al abordar el tema es la notación ya que muchas de las nociones introducidas en el tema, así como su significado y utilización son nuevas para los alumnos. Por ejemplo, el símbolo de sumatorio Σ y su significado o la notación para designar la desviación típica σ . Es la primera vez que ven este símbolo, por lo que hay que escribirlo en la pizarra, junto con su nombre 'sigma' y decir a los alumnos que lo apunten para denotarlo correctamente.

Una dificultad inherente al tema es la diferenciación entre relación estadística y relación funcional.

Interpretar correctamente el significado del valor numérico del coeficiente de correlación es una dificultad para indicar la fiabilidad de las estimaciones.

Otra de las dificultades previstas es relacionar la interpretación gráfica de la nube de puntos con la correlación entre las variables en función de dos aspectos, correlación fuerte o débil y correlación positiva o negativa. Y relacionar esta interpretación con los resultados obtenidos por medios analíticos.

6.2. Errores y su posible origen

Un error previsible es no asimilar el concepto de valor absoluto del coeficiente de correlación, y relacionar el valor positivo del coeficiente con una correlación fuerte y el valor negativo o decreciente de la nube de puntos con una correlación débil. Esto puede deberse a relacionar el coeficiente de correlación y su interpretación con la recta de los números reales.

Debido a errores de cálculo se puede llegar a valores del coeficiente de correlación mayores que 1 o menores que -1, sin tener en cuenta el rango en el que se encuentra este coeficiente ($-1 < r < 1$). De esta manera se asocian valores altos a una correlación fuerte y valores bajos a una correlación débil. Este error también se debe a la interpretación del coeficiente de correlación en relación con la recta real.

7. El proceso de estudio

En este capítulo se detallan las sesiones de práctica docente realizadas durante la estancia en el centro para la impartición del tema de Estadística Bidimensional.

7.1. Distribución del tiempo de la clase

El tema se planifica para impartirlo en 5 sesiones de docencia más una sesión de prueba corta, puntuable para la nota final. Se describe la distribución del tiempo de la clase en sesiones de 55 minutos.

Tabla 43

Sesión 1			
Tipo	Tiempo	Responsable	Tipo de docencia
Introducción del tema	10'	Compartida	Dialógica
Teoría 1-Distribución bidimensional	10'	Profesor	Magistral
Teoría 2-Nube de puntos	5'	Profesor	Dialógica
Teoría 3-Correlación	10'	Profesor	Dialógica
Teoría 4- Parámetros estadísticos	15'	Compartida	Dialógica
Tabla	5'	Profesor	Dialógica

Tabla 44

Sesión 2			
Tipo	Tiempo	Responsable	Tipo de docencia
Repaso sesión 1	5'	Compartida	Dialógica
Tabla	10'	Compartida	Dialógica
Ejemplo	15'	Compartida	Constructivista
Ejercicio 1	20'	Alumnos	Constructivista
Explicación tarea	5'	Profesor	Magistral

Tabla 45

Sesión 3			
Tipo	Tiempo	Responsable	Tipo de docencia
Repaso sesión 2	5'	Compartida	Dialógica
Corrección tarea	20'	Compartida	Dialógica
Teoría 5- Coeficiente de correlación lineal	15'	Profesor	Magistral
Teoría 6- Recta de regresión de Y sobre X	15'	Profesor	Dialógica

Tabla 46

Sesión 4			
Tipo	Tiempo	Responsable	Tipo de docencia
Repaso sesión 3	5'	Compartida	Dialógica
Ejemplo (Coeficiente de correlación y recta de regresión)	10'	Compartida	Dialógica
Teoría: Estimación y fiabilidad.	15'	Profesor	Dialógica
Ejercicio 1	20'	Alumnos	Constructivista
Explicación tarea: Ejercicio 2	5'	Profesor	Dialógica

Tabla 47

Sesión 5			
Tipo	Tiempo	Responsable	Tipo de docencia
Repaso sesión 4	5'	Compartida	Dialógica
Corrección tarea	20'	Alumnos	Dialógica
Ejercicio 4	25'	Alumnos	Constructivista
Repaso	10'	Compartida	Dialógica

7.2. Actividades adicionales planificadas

Las actividades adicionales que se planifican son ejercicios y problemas de distribuciones bidimensionales que se encuentran en el Anexo C de esta memoria.

7.3. La tarea: actividad autónoma del alumno prevista

En este apartado se adjuntan las tablas con la actividad autónoma prevista por los alumnos para realizar la tarea mandada y reforzar los conceptos vistos en clase.

Tabla 48

Tipo	Tiempo estimado	Relación con el proceso de enseñanza aprendizaje
Sesión 1		
Repaso teoría clase	20'	Refuerzo
Sesión 2		
Estudio personal	15'	Refuerzo
Actividades	30'	Aplicación
Sesión 3		
Estudio personal	20'	Refuerzo
Sesión 4		
Estudio personal	10'	Refuerzo
Actividades	40'	Aplicación
Sesión 5		
Actividades	30'	Aplicación
Estudio personal	30'	Refuerzo

8. Experimentación

Este capítulo se basa en la experimentación que se lleva a cabo en la clase de 4º de ESO del centro de prácticas.

Para ello se define la muestra y se analiza el diseño de la prueba que se realiza a los alumnos. Después se analizan los comportamientos esperados por los alumnos y los resultados obtenidos, para discutir finalmente si los resultados obtenidos coinciden con los esperados.

8.1. Método

La evolución de una teoría en didáctica de las matemáticas puede determinarse por el contraste entre un análisis *a priori* y un análisis *a posteriori*. La teoría busca validar las hipótesis que formula (*a priori*). Los hechos observados permiten (*a posteriori*) validar o refutar, total o parcialmente, las hipótesis enunciadas.

La *ingeniería didáctica* (Artigue, 1989) permite abordar el contraste experimental necesario, que permita determinar condiciones de *reproducibilidad* de situaciones didácticas. Aquí, las *variables didácticas* actúan de “contraste o reactivo” que permiten de manera controlada provocar en los sujetos modificaciones en sus estrategias de acción para adaptarlas al medio.

El estudio de la adecuación de las variables didácticas para determinar cambios en las estrategias de acción representa un instrumento de validación interna de las conclusiones que puedan extraerse de una observación concreta. En estas condiciones, se puede definir una situación *reproducible*; es decir, en condiciones similares, con un control del medio, la construcción del conocimiento pretendido será la misma.

La cuestión de la reproducibilidad de las situaciones incide sobre la fiabilidad de las observaciones y, sobre todo, sobre su validez. La fiabilidad presupone una estabilidad en el funcionamiento del sistema didáctico; el contraste repetido entre el análisis *a priori* y el análisis *a posteriori* permite hacer evolucionar las condiciones del medio (incluidas las intervenciones del profesor) que garantizan la construcción del saber pretendido, de tal manera que la situación devenga reproducible. Es entonces cuando su *validez* puede ser aceptada, puesto que la situación es exitosa y aplicable de manera estable.

En este trabajo, la parte I “La estadística en el currículo vigente y en los libros de texto” constituye el estudio previo de la dimensión de enseñanza, desde una perspectiva eminentemente institucional; a saber:

1. El contenido matemático en el currículo vigente, incluidas las orientaciones y criterios de evaluación.
2. El desarrollo de estas directrices oficiales en los libros de texto escolares.

Este estudio precede al análisis *a priori* realizado en los capítulos 5, 6 y 7, donde se abordan las dimensiones:

- *Epistemológica*: las matemáticas presentes en la unidad didáctica objeto de estudio.

- *Cognitiva*: dificultades y errores de los estudiantes en el aprendizaje de la unidad didáctica.
- *De enseñanza*: descripción del proceso de estudio implementado.

En el capítulo 8, este análisis *a priori* es contrastado con los resultados de la experimentación, permitiendo una valoración de los mismos basada en las “expectativas previas” (*discusión de los resultados*), que supone la fase última del método de la ingeniería didáctica.

8.2. Muestra y diseño de la experimentación

La muestra que se estudia corresponde a una clase de 4º de ESO del colegio Ntra. Sra. del Puy (Estella) de 21 alumnos.

Se trata de un grupo homogéneo, perteneciente en su mayoría a un estrato social de clase media. En el grupo hay dos alumnas extranjeras de nacimiento, su integración en clase es correcta y su lengua materna es el castellano, por lo que el nivel de asimilación de conocimientos por estos motivos no está en desventaja respecto a los alumnos españoles.

Cabe mencionar que la clase en la que se realiza la experimentación es un curso de 4º de ESO encaminado a hacer Bachillerato, por lo que la capacidad y motivación del alumnado respecto a la asignatura es superior a los alumnos de otras clases que van dirigidas a Formación Profesional o alumnos con dificultades para obtener el título de ESO.

El objetivo principal de la experimentación es observar las características que muestran los alumnos a la hora de resolver problemas de distribuciones bidimensionales.

La programación de aula para este tema, no preveía un examen específico, sino que este se hacía en conjunto con el tema siguiente. Por este motivo, se realiza una ‘prueba corta’, de 40 minutos de duración que no puntúa como un examen. Sin embargo, para que los alumnos preparen la prueba y los resultados obtenidos puedan ser analizados, la prueba computa como un ejercicio puntuable para el cálculo de la nota trimestral de la asignatura.

Por ello, el nivel de dificultad y la extensión de la prueba se limitan por el departamento, para que se adapte a la programación de la asignatura.

8.3. El cuestionario

El cuestionario pretende estudiar el nivel de asimilación de los conocimientos de distribuciones bidimensionales que han adquirido los alumnos a lo largo de las cinco sesiones que se han trabajado en clase y mediante la tarea autónoma prevista.

Dado que el tiempo ha sido limitado y el tema corresponde a un curso superior, el cuestionario planteado es de una dificultad media. El objetivo es que la mayoría de los alumnos realicen los problemas planteados, y para ello se ha diseñado un cuestionario con dos problemas.

En el primero tienen que hallar los parámetros estadísticos de la distribución bidimensional, que aunque no se piden explícitamente son necesarios para hallar posteriormente la recta de regresión y realizar la estimación.

Se trata de un problema en el que el coeficiente de correlación es muy alto $r = 0.965$, próximo a 1, para que la estimación sea fiable.

Problema 1

Se ha analizado en distintos modelos de impresoras cuál es el coste por página (en céntimos de euro) en blanco y negro y cuál es el coste por página si esta es en color. La siguiente tabla nos da los seis primeros pares de datos obtenidos:

X: B Y N	8	11	17	21	14	10
Y: Color	33	49	95	106	58	53

- Halla la recta de regresión de Y sobre X.
- ¿Cuánto nos costaría imprimir una página en color en una impresora en la que el coste por página en blanco y negro fuera de 12 céntimos de euro? ¿Es fiable la estimación? Razona la respuesta.

Ilustración 16

Dado que el tiempo de realización de la prueba es de 40 minutos, y en el primer ejercicio los alumnos han realizado la tabla, en el segundo se dan los sumatorios de los valores que necesitan para calcular los parámetros estadísticos.

De esta manera se les pide hallar la covarianza y el coeficiente de correlación en el apartado a, para lo que tienen que calcular antes la media y las desviaciones típicas de las variables.

En este problema el coeficiente de correlación es muy bajo $r = 0.097$, próximo a 0, por lo que no hay relación entre las dos variables.

Problema 2

Se han realizado unas pruebas de habilidad (puntuán de 0 a 5) en un grupo de alumnos. Las siguientes puntuaciones corresponden a las obtenidas por varios alumnos en dos de ellas:

1ª Prueba	5	5	4	3	2	4
2ª Prueba	4	3	4	4	3	2

Se han obtenido los siguientes resultados para los datos anteriores:

$\sum x_i = 23$	$\sum y_i = 20$	$\sum x_i^2 = 95$	$\sum y_i^2 = 70$	$\sum x_i y_i = 77$
-----------------	-----------------	-------------------	-------------------	---------------------

- Calcula la covarianza y el coeficiente de correlación.
- Explica razonadamente cómo es la relación entre las dos variables.

Ilustración 17

Criterios de corrección

En el centro hay cuatro grupos de 4º de ESO, en los que tienen docencia de matemáticas tres profesoras diferentes. Por ello, tras realizar el examen, las profesoras ponen los criterios de corrección comunes para que la corrección de los mismos sea lo más unificada posible y no haya diferencias significativas.

Por una parte tenemos las puntuaciones de cada apartado de los problemas, en este caso el primero puntúa sobre 18. Y por otra parte, los errores o fallos que restan puntuación como por ejemplo, no redondear bien o no poner las fórmulas.

Tabla 49

Problema 1		
Variable		Nota (_/18)
Tabla		1
Parámetros	Media	2
	Varianza	2
	Desviación típica	2
	Covarianza	1
Recta de regresión	Fórmula	2
	Recta	2
Estimación		2
Fiabilidad		1
Razonamiento		2
Coeficiente de correlación		1
Aspectos que restan puntos	Error de cálculo sencillo	-0.2
	Error de cálculo grave	No puntúa
	No poner la fórmula	-0.2
	Fórmula incorrecta	No puntúa
	No detallar el nombre de los parámetros	-0.1
	No redondear bien	-0.1

El segundo ejercicio puntúa sobre 9, ya que el tiempo de realización es menor y los cálculos que hay que realizar también son menores. Se sigue el mismo criterio que en el anterior para

restar puntuación si los alumnos no redondean correctamente o si no escriben bien las fórmulas a aplicar en el ejercicio.

Tabla 50

Problema 2		
Variable		Nota (_/9)
Parámetros	Medias	2
	Covarianza	1
	Desviaciones típicas	2
	Coeficiente de correlación	1
Relación entre las variables	Fuerte / débil	1
	Razonamiento	1
	Positiva / negativa	1
Aspectos que restan puntos	Error de cálculo sencillo	-0.2
	Error de cálculo grave	No puntúa
	No poner la fórmula	-0.2
	Fórmula incorrecta	No puntúa
	No detallar el nombre de los parámetros	-0.1
	No redondear bien	-0.1

8.4. Cuestiones y comportamientos esperados

Como se ha comentado anteriormente la organización del curso prevé para el tema de Distribuciones bidimensionales 5 sesiones. El tiempo es ajustado para explicar en profundidad un tema que es nuevo para los alumnos. Por ello, no se esperan grandes resultados en la realización de las pruebas.

Se pretende que los alumnos sean capaces de realizar los cálculos mecánicamente para llegar a realizar estimaciones. El cálculo mecánico de los parámetros estadísticos sería un objetivo mínimo que la mayoría de los alumnos deberían conseguir, así como la correcta realización de la tabla con los sumatorios correspondientes. Del mismo modo, la realización de la nube de puntos sería un objetivo mínimo a realizar por la mayor parte del alumnado.

Hallar la covarianza o la correlación entre las dos variables serían objetivos medios, que podrían alcanzar la media de los alumnos, pero quedando ya una parte de ellos por debajo de este objetivo, ya que son nuevos conceptos que es la primera vez que ven. También serían

objetivos medios, la interpretación de la nube de puntos y la comprobación de existencia de relación lineal entre las variables a estudiar por medio de esta.

Por último, realizar correctamente las estimaciones, y sobre todo, razonar correctamente la fiabilidad de estas sería un objetivo máximo, que realizarían una minoría de los alumnos.

Para los problemas planteados en la prueba se establecen los objetivos mínimos, medios y máximos previsibles a alcanzar por los alumnos.

Tabla 51

Problema 1	
Objetivos mínimos	Realización de la tabla con los sumatorios
	Cálculo de los parámetros estadísticos
Objetivos medios	Recta de regresión
Objetivos máximos	Estimación
	Fiabilidad de la estimación y razonamiento

Tabla 52

Problema 2	
Objetivos mínimos	Cálculo de los parámetros
Objetivos medios	Coeficiente de correlación
Objetivos máximos	Razonamiento de la relación entre las variables

8.5. Resultados

Tras analizar el examen y de acuerdo a los objetivos esperados y a los criterios de corrección se determinan las variables de estudio de los resultados obtenidos.

Problema 1:

- V 1.1: Realización de la tabla
- V 1.2: Cálculo de parámetros
- V 1.3: Recta de regresión
- V 1.4: Estimación
- V 1.5: Coeficiente de correlación
- V 1.6: Razonamiento (fiabilidad de la estimación)

Problema 2:

- V 2.1: Medias
- V 2.2: Desviaciones típicas

- V 2.3: Covarianza
- V 2.4: Coeficiente de correlación
- V 2.5: Explicación correlación: fuerte/débil
- V 2.6: Explicación correlación: positiva/negativa

Otras variables:

- V 3.1: Redondeo correcto
- V 3.2: Fórmulas correctas

Fallos:

- F 4.1: Estimación por regla de tres.
- F 4.2: Confunde términos (coef. correlación/covarianza) o (varianza/desviación típica)

En las Tabla 53 y Tabla 54 se muestran los resultados obtenidos por los 21 alumnos de la clase de 4º para las variables analizadas. Se estudia los alumnos que realizan correctamente cada variable (X), los que tienen algún error en el cálculo o la realización (E) y los que no realizan ese apartado (-). Por otra parte también se han tenido en cuenta las variables que restan puntuación como es redondear correctamente o escribir las fórmulas.

Tras realizar la prueba se han visto dos tipos de fallos que no se habían prevenido como son realizar la estimación mediante una regla de tres y la confusión de los términos tratados, como por ejemplo, confundir la covarianza con el coeficiente de correlación o la varianza con la desviación típica.

Tabla 53

Alumno (Nota total)	Problema 1									
	V 1.1	V 1.2	V 1.3	V 1.4	V 1.5	V 1.6	V 3.1	V 3.2	F 4.1	Nota (_/18)
1 (7.9)	X	X	X	X	X	X	-	X	-	13.7
2 (2.3)	E	X	-	-	E	-	-	E	-	3.5
3 (6.1)	X	X	-	-	X	-	X	X	-	8.9
4 (6.5)	E	X	E	-	X	X	-	X	-	11.6
5 (0.4)	X	-	-	-	-	-	-	E	-	1.0
6 (8.9)	X	X	X	X	X	X	X	X	-	17.9
7 (7.5)	E	X	X	X	X	E	-	X	-	14.7
8 (9.6)	X	X	X	X	X	X	X	X	-	17.9
9 (4.0)	E	E	-	-	-	-	-	X	-	5.3
10 (9.5)	X	X	X	X	X	X	X	X	-	17.9
11 (9.9)	X	X	X	X	X	X	X	X	-	17.9
12 (4.0)	X	X	E	E	-	E	-	X	-	7.8
13 (5.3)	E	X	E	-	-	-	-	X	-	9.4
14 (1.8)	X	-	-	-	-	-	-	-	X	1.0
15 (8.7)	X	X	X	E	X	X	X	X	-	17.6
16 (2.4)	X	-	-	-	-	-	-	-	X	1.0
17 (7.3)	X	E	E	-	X	-	-	X	-	11.8
18 (9.5)	X	X	E	X	X	X	X	X	-	17.7
19 (2.0)	-	-	-	-	-	-	-	-	X	0
20 (8.0)	E	X	X	X	X	X	-	X	-	13.9
21 (8.4)	X	X	X	X	X	X	-	X	-	17.7

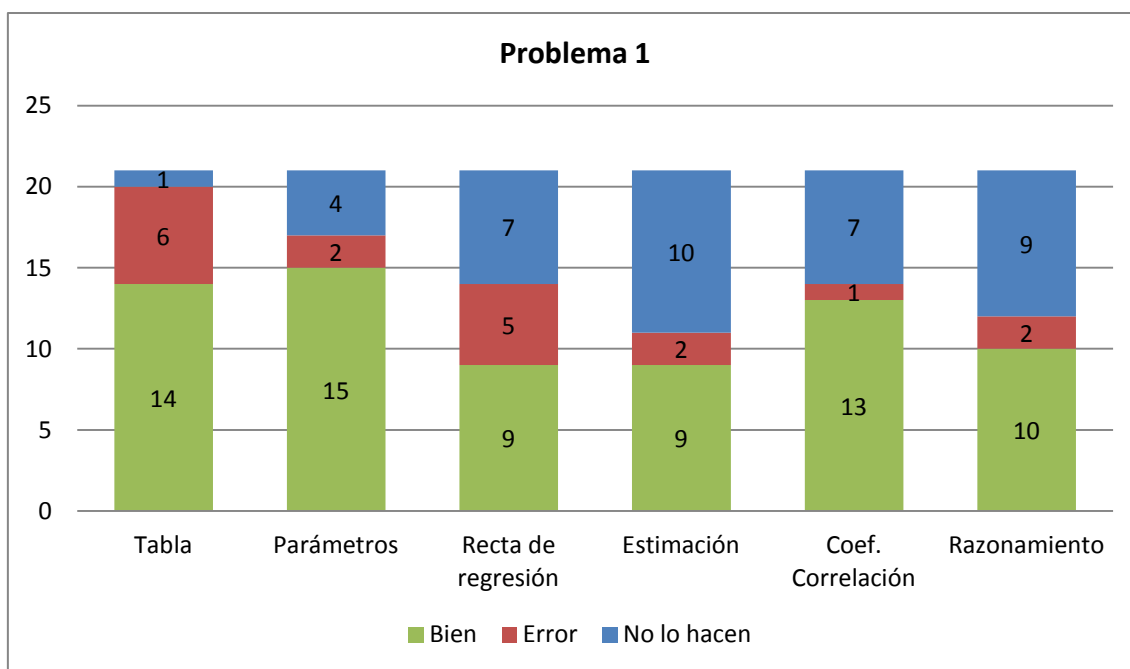


Gráfico 1

Tabla 54

Alumno (Nota total)	Problema 2									
	V 2.1	V 2.2	V 2.3	V 2.4	V 2.5	V 2.6	V 3.1	V 3.2	F 4.2	Nota (_/9)
1 (7.9)	X	X	X	X	X	-	-	X	-	7.6
2 (2.3)	E	-	E	-	E	-	-	E	-	2.7
3 (6.1)	X	X	E	X	X	-	X	X	-	7.7
4 (6.5)	X	X	X	X	-	-	-	X	-	5.9
5 (0.4)	-	-	-	-	-	-	-	E	-	0
6 (8.9)	X	X	X	X	E	-	X	X	-	6.0
7 (7.5)	X	E	X	X	E	-	-	X	-	5.5
8 (9.6)	X	X	X	X	X	-	X	X	-	7.9
9 (4.0)	X	E	-	E	-	-	-	X	-	5.5
10 (9.5)	X	X	E	X	X	-	X	X	-	7.8
11 (9.9)	X	X	X	E	X	X	X	X	-	8.8
12 (4.0)	X	-	X	E	E	-	-	X	-	2.9
13 (5.3)	X	X	X	X	-	-	-	X	-	4.9
14 (1.8)	X	X	X	-	E	-	-	-	X	3.9
15 (8.7)	X	X	X	X	E	E	X	X	-	6.0
16 (2.4)	X	E	X	X	E	-	-	-	X	5.6
17 (7.3)	X	X	X	E	X	-	-	X	-	7.8
18 (9.5)	X	X	X	X	X	-	X	X	-	8.0
19 (2.0)	X	X	E	E	-	-	-	-	-	5.5
20 (8.0)	X	X	X	E	X	-	-	X	-	7.8
21 (8.4)	X	E	X	X	E	-	-	X	-	4.9

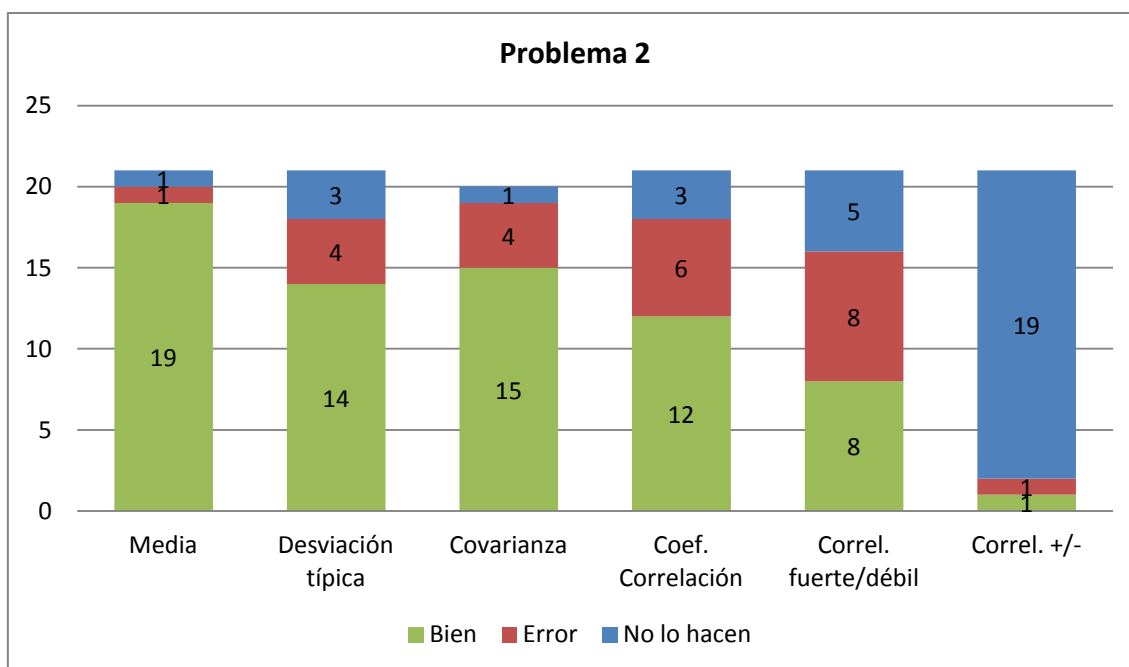


Gráfico 2

8.6. Discusión de los resultados

Tras analizar los resultados obtenidos vemos que la mayoría de los alumnos ha asimilado correctamente las principales nociones explicadas en el tema. Los resultados obtenidos están en relación con las cuestiones y comportamientos esperados que hemos analizado en el apartado cuarto del presente capítulo.

Como vemos en el Gráfico 1, en el primer ejercicio la mayoría de los alumnos ha realizado correctamente la tabla y el cálculo de los parámetros estadísticos necesarios para la resolución del primer apartado. A la hora de calcular la recta de regresión el porcentaje de alumnos que la realizan correctamente pasa a ser menos de la mitad. Llama la atención que esto se deba a que tienen muchos errores de cálculo en la fórmula de la recta de regresión a la hora de utilizar procedimientos como desarrollar un paréntesis o aplicar denominador común.

En el segundo apartado del problema (Ilustración 16) se pide realizar una estimación y razonar la fiabilidad de la misma. El cálculo de la estimación lo realizan bien los alumnos que han hallado correctamente la recta de regresión del apartado anterior, ya que se trata de una sustitución en la ecuación que no plantea ninguna dificultad.

Llama la atención que entre los alumnos que realizan mal la estimación hay tres que la realizan por regla de tres. Toman dos elementos de la tabla planteada en el problema y realizan la estimación pedida a partir de estos. En principio es un comportamiento que llama la atención, ya que precisamente el tema de distribuciones bidimensionales trata de estudiar si hay relación entre dos variables dadas y cómo es ésta. Sin embargo, se explica la utilización de este procedimiento debido al abuso de utilización de la linealidad (Van Dooren, De Bock, & Verschaffel, 2006).

El coeficiente de correlación lo calculan bien más de la mitad de los alumnos, y a partir de este casi llega a la mitad el número de alumnos que explican razonadamente cómo es la fiabilidad de las variables. Siendo éste un objetivo máximo se cumplen las expectativas planteadas en los comportamientos esperados.

Para el segundo problema (Ilustración 17) se da a los alumnos los sumatorios de la tabla, ya que disponen de poco tiempo, y es un trabajo mecánico que han demostrado en el primer problema si saben hacerlo. A pesar de ello, es desconcertante el hecho de que cinco alumnos vuelven a realizar la tabla, con la consiguiente pérdida de tiempo que les conlleva.

En el primer apartado se pide calcular la covarianza y el coeficiente de correlación, para lo que implícitamente tienen que calcular las medias y las desviaciones típicas de las variables. Vemos en el Gráfico 2 que casi la totalidad de los alumnos calcula las medias correctamente, se trata de un objetivo mínimo que deben realizar correctamente, ya que es una noción comienzan a estudiar en el tercer ciclo de primaria (Tabla 1). En el cálculo de las desviaciones típicas, covarianza y coeficiente de correlación aumenta el número de alumnos que fallan, principalmente debido a errores de cálculo.

En el segundo apartado tienen que explicar cómo es la relación entre las variables a partir de los datos obtenidos anteriormente. En este caso el número de alumnos que explica correctamente la relación baja a 8, mientras que hay otros 8 que explican mal esta relación. Parte de los que explican mal la relación cometen un error de aproximación del número del coeficiente de correlación. El resultado de éste es bajo $r = 0.097$, ($0 < r < 0.3$), por lo que la relación entre las variables es débil (Ilustración 15). Sin embargo cuatro de estos alumnos dicen que “está próximo a 1”, por lo que consideran que la relación entre las variables es fuerte y la estimación fiable. Dado el curso en el que se encuentran estos alumnos y consideradas sus capacidades, se concluye que puede ser debido a un despiste más que a una mala comprensión de los números reales.

Por otra parte, sólo dos alumnos explican si la relación entre las variables es positiva o negativa. Se demuestra el hecho de que si no se pide explícitamente a los alumnos las cuestiones a desarrollar, no añaden información que puede ser útil para la explicación del problema.

En relación a esto, también llama la atención que solamente una alumna dibuja la nube de puntos, ya que ésta no es pedida explícitamente en el problema, ni puntúa para la nota. Vemos que hay procedimientos como la representación gráfica, que ayudan a los alumnos a una mejor comprensión del problema o a comprobar los resultados obtenidos, y que sin embargo, si no se piden explícitamente, aunque mejoren la explicación del problema, no son realizados.

9. Síntesis, conclusiones y cuestiones abiertas

Síntesis

En este Trabajo Fin de Máster se ha pretendido hacer un análisis del tema de Distribuciones bidimensionales en 4º de ESO, dentro de lo que el contexto y habilidad de la autora han permitido.

Para ello, en la primera parte se ha analizado el estado de la legislación actual exigida por el Ministerio de Educación en relación con el tema impartido. También se ha analizado cómo los libros de texto utilizados en el centro plasman esta legislación y cómo desarrollan el tema estudiado desde una perspectiva longitudinal del currículo.

En la segunda parte del presente trabajo, se ha analizado la experiencia docente llevada a cabo en el centro de prácticas. Se ha estudiado la impartición del tema de Distribuciones bidimensionales y el análisis de la prueba que realizaron los alumnos con las conclusiones que se han extraído.

Conclusiones

La primera conclusión que se extrae es relativa a la organización del currículo en el bloque de Estadística por parte del Ministerio de Educación.

Los contenidos de estadística a lo largo de los cuatro cursos de la ESO son muy similares, prácticamente no se introducen nuevas nociones, sino que se vuelve a revisar lo dado en cursos anteriores desde una perspectiva superior o con ejercicios más complicados.

Debido a esto, en la organización del bloque de estadística en Bachillerato ocurre lo contrario. Se introduce gran cantidad de nuevas nociones y procedimientos en dos cursos, lo que resulta difícil de asimilar y trabajar correctamente por parte de los alumnos. Además en la rama de Ciencias, aunque aparece el tema de Distribuciones bidimensionales, los centros no lo dan, ya que no entra en selectividad y tienen muchísima materia para desarrollar en la asignatura de matemáticas.

Por lo tanto, la decisión del centro de adelantar el tema de Distribuciones bidimensionales a 4º de ESO me parece acertada y conveniente. Debido al carácter obligatorio de este curso, puede haber alumnos a los que les cueste asimilar ciertos conceptos, pero la mayoría de los alumnos tienen capacidad suficiente para poder trabajar correctamente el tema. Aunque la mayoría de ellos lo traten de una manera mecánica y superficial, es una manera de que vayan familiarizándose con las nuevas nociones y cuando vuelvan a verlas en cursos superiores tengan unos conocimientos básicos de la estadística bidimensional. Además, los alumnos más avanzados sí que son capaces de asimilar correctamente las nociones explicadas y las sutilezas y particularidades del tema.

Otra conclusión ha sido que las nociones matemáticas, por sencillas que sean, se entienden mejor mediante ejemplos. Esto puede ser debido al carácter abstracto de las nociones y

definiciones matemáticas, a las que los alumnos no están acostumbrados y asimilan mejor a base de ejemplos. De la misma manera, estos nuevos conceptos se consolidan a base de ejercicios. No basta con que los alumnos entiendan el concepto, sino que deben interiorizarlo y afianzarlo a través de trabajo autónomo mediante ejercicios y problemas.

Por último, es reseñable que la mayoría de los errores producidos por los alumnos no son debidos a la comprensión de las nociones tratadas en el tema, estas las asimilan con facilidad, sino principalmente a errores de cálculo en el desarrollo de los ejercicios. Tienen dificultades con el uso del álgebra y las propiedades de las operaciones matemáticas básicas como son el orden de operaciones, las factorizaciones o los denominadores comunes.

Tras el análisis de la práctica docente realizada en el centro de prácticas, se detectan las carencias observadas de los resultados analizados. Si se volviese a impartir el tema, se modificarían o introducirían aspectos diferentes para conseguir una mejor calidad de la docencia del tema dado. En primer lugar, se profundizaría en la diferencia entre relación funcional (lineal, cuadrática, ...) y estadística, para evitar el error de realizar estimaciones mediante regla de tres. También se insistiría más en la importancia y ayuda de la representación gráfica para la realización y comprobación de los ejercicios. Por último, se explicaría el sentido del coeficiente de correlación, indicador de la calidad del ajuste, como valor absoluto, para evitar la confusión del signo de la pendiente de la recta con un buen o mal ajuste.

Cuestiones abiertas

La principal cuestión que se plantea es la conveniencia de impartir el tema de Distribuciones bidimensionales en 4º de ESO en vez de en 1º de Bachillerato. Dadas las características del Prácticum y el contexto en el que se ha realizado, la muestra realizada no es suficientemente representativa para extrapolar los resultados al resto de la población. Sería conveniente realizar un estudio con una muestra representativa de la impartición de este tema en 4º y en 1º, para analizar las diferencias, resultados y conclusiones a las que se llega en ambos cursos.

Otra cuestión es la relativa a la conveniencia de utilización de las TIC's. Por una parte son una gran ventaja, facilitan el trabajo mecánico que tienen que desarrollar los alumnos, tienen una gran potencialidad a la hora de trabajar con grandes cantidades de datos y presentan la información y los resultados de inmediato. Pero por otra, en cursos en los que los alumnos no han desarrollado todavía la práctica y el manejo correcto del álgebra, limitan su práctica, todavía necesaria en estos niveles y que se trata en el resto de la asignatura como contenidos transversales.

Por último, se plantea la mejor manera de impartir este tema para conseguir una mejor asimilación de los conocimientos. Dado el nivel en el que estamos, podría desvincularse del formalismo matemático, que impide a muchos alumnos alcanzar los conocimientos propuestos, que si consiguen entender a base de ejemplos cercanos.

Referencias

- Artigue, M. (1989). Ingénierie didactique. *Recherches en Didactique des Mathématiques*, 9(3), 282-307.
- Ciencia, M. d. (8 de diciembre de 2006). Real Decreto. págs. 233-244.
- Colera, J., & Gaztelu, I. (2008). *Matemáticas 2º ESO*. Madrid: Anaya.
- Colera, J., & Oliveira, M. (2009). *Matemáticas aplicadas a las Ciencias Sociales II. Bachillerato*. Madrid: Anaya.
- Colera, J., García, R., Oliveira, M., & Santanella, E. (2008). *Matemáticas aplicadas a las Ciencias Sociales I. Bachillerato*. Madrid: Anaya.
- Colera, J., García, R., Oliveira, M., & Santanella, E. (2008). *Matemáticas I. Bachillerato*. Madrid: Anaya.
- Colera, J., Martínez, M., Gaztelu, I., & Oliveira, M. (2008). *Matemáticas 3º ESO*. Madrid: Anaya.
- Colera, J., Martínez, M., Gaztelu, I., & Oliveira, M. (2008). *Matemáticas 4º ESO (Opción B)*. Madrid: Anaya.
- Godino, J., Font, V., & Wilhelmi, M. (2006). Análisis ontosemiótico de una lección sobre la suma y la resta. *Revista Latinoamericana de Investigación en Matemática Educativa*, 133-156.
- Ministerio de Educación y Ciencia (2006). *Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación primaria*, (BOE 293, de 8 de diciembre, 43053-43102).
- Ministerio de Educación y Ciencia (2007). *Real Decreto 1631/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Secundaria Obligatoria*, (BOE 5, de 5 de enero, 677-773).
- Ministerio de Educación y Ciencia (2007). *Real Decreto 1467/2007, de 2 de noviembre, por el que se establece la estructura del bachillerato y se fijan sus enseñanzas mínimas*, (BOE 266, de 6 de noviembre, 45381-45477).
- Spiegel, M. R., & Stephens, L. J. (2009). *Estadística*. México: Mc Graw Hill.
- Van Dooren, W., De Bock, D., & Verschaffel, L. (2006). La búsqueda de las raíces de la ilusión de linealidad. *Boletín de Estudios e Investigación, Monografía IV*, 115-135.

Referencia de Gráficos

Gráfico 1 81

Gráfico 2 83

Referencia de Ilustraciones

Ilustración 1..... 56

Ilustración 2..... 57

Ilustración 3..... 58

Ilustración 4..... 58

Ilustración 5..... 58

Ilustración 6..... 59

Ilustración 7..... 59

Ilustración 8..... 60

Ilustración 9..... 60

Ilustración 10..... 61

Ilustración 11..... 62

Ilustración 12..... 64

Ilustración 13..... 64

Ilustración 14..... 65

Ilustración 15..... 66

Ilustración 16..... 75

Ilustración 17..... 75

Referencia de Tablas

Tabla 1 12

Tabla 2 14

Tabla 3 15

Tabla 4 17

Tabla 5 17

Tabla 6 20

Tabla 7 21

Tabla 8 22

Tabla 9 23

Tabla 10	23
Tabla 11	25
Tabla 12	25
Tabla 13	28
Tabla 14	28
Tabla 15	30
Tabla 16	31
Tabla 17	31
Tabla 18	33
Tabla 19	33
Tabla 20	34
Tabla 21	34
Tabla 22	36
Tabla 23	36
Tabla 24	37
Tabla 25	37
Tabla 26	38
Tabla 27	38
Tabla 28	39
Tabla 29	39
Tabla 30	42
Tabla 31	42
Tabla 32	43
Tabla 33	43
Tabla 34	44
Tabla 35	45
Tabla 36	53
Tabla 37	53
Tabla 38	54
Tabla 39	54
Tabla 40	54
Tabla 41	55
Tabla 42	57

Tabla 43	69
Tabla 44	69
Tabla 45	70
Tabla 46	70
Tabla 47	70
Tabla 48	71
Tabla 49	76
Tabla 50	77
Tabla 51	78
Tabla 52	78
Tabla 53	80
Tabla 54	82

Anexos

A. Unidad didáctica del libro de texto



V ESTADÍSTICA Y PROBABILIDAD

El pensamiento determinístico y el pensamiento aleatorio juegan sus papeles, unas veces como adversarios y otras como colaboradores, en el desarrollo de la ciencia.

"La estadística es la única herramienta con la que puede practicarse una brecha a través de la formidable espesura de dificultades que obstruyen el sendero de quienes buscan la ciencia del hombre".

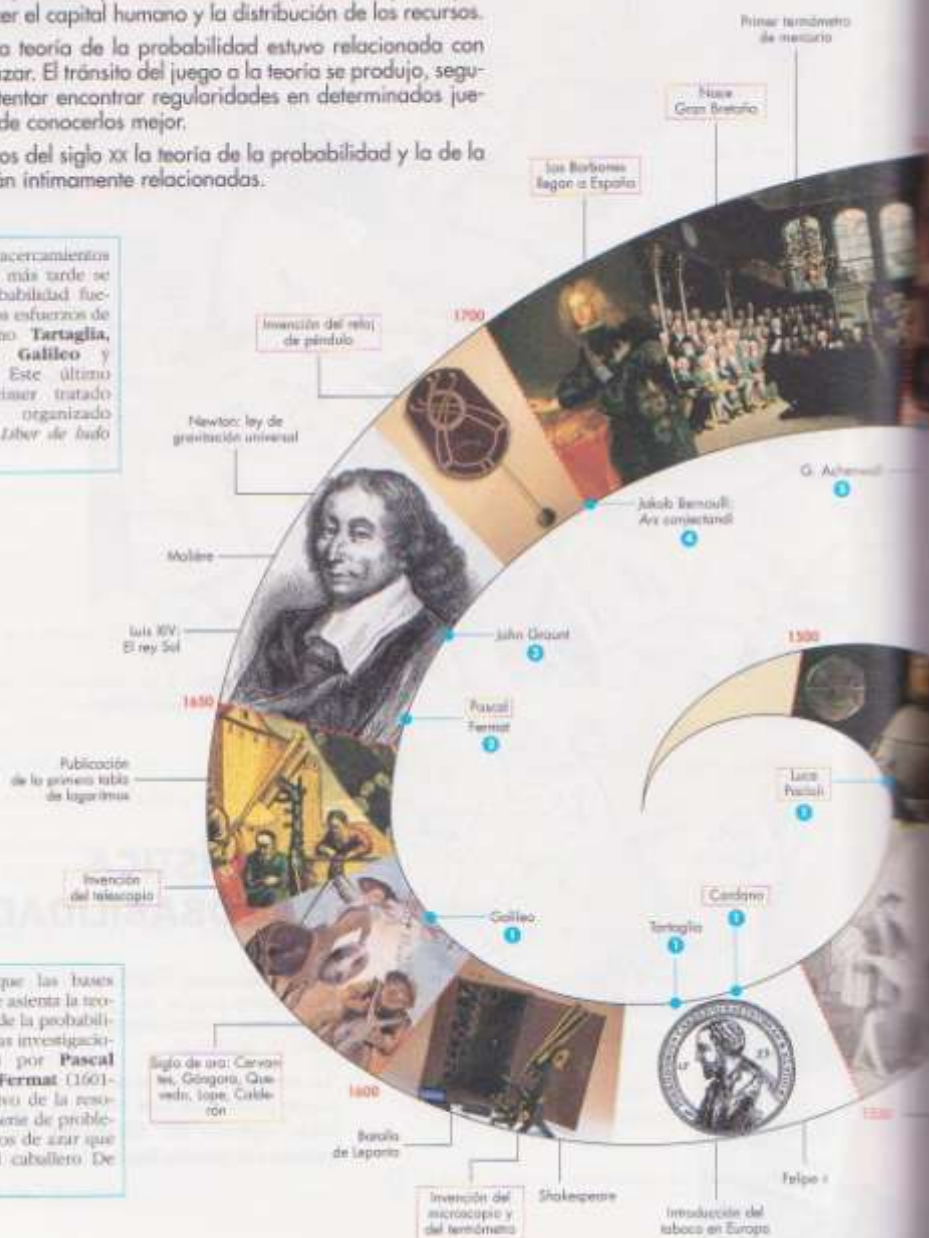
Francis Galton (1822-1911)

NOTAS HISTÓRICAS. ESTADÍSTICA Y PROBABILIDAD

El origen de la estadística está estrechamente relacionado con los censos realizados a lo largo de la historia. Desde las culturas más antiguas (China, Roma...) hasta nuestros días, existe la preocupación por conocer el capital humano y la distribución de los recursos. En principio, la teoría de la probabilidad estuvo relacionada con los juegos de azar. El tránsito del juego a la teoría se produjo, seguramente, al intentar encontrar regularidades en determinados juegos con el fin de conocerlos mejor. Desde principios del siglo XX la teoría de la probabilidad y la de la estadística están íntimamente relacionadas.

1 Los primeros acercamientos serios a lo que más tarde se llamaría la probabilidad fueron debidos a los esfuerzos de personajes como **Tartaglia**, **Luca Pacioli**, **Galileo** y **G. Cardano**. Este último escribe el primer tratado medianamente organizado sobre el azar: *Liber de ludo alea*.

2 Parece claro que las bases sobre las que se asienta la teoría matemática de la probabilidad parten de las investigaciones realizadas por **Pascal** (1623-1662) y **Fermat** (1601-1665) con motivo de la resolución de una serie de problemas sobre juegos de azar que les propuso el caballero De Méré.





13

DISTRIBUCIONES BIDIMENSIONALES

El término regresión fue acuñado por **sir Francis Galton** (1822-1911), primo del famoso naturalista **Darwin**. Galton estudió la siguiente distribución bidimensional:

Estatura media de un matrimonio - estatura media de sus hijos adultos

Encontró una correlación fuerte: cuanto mayor era la primera, mayor era la segunda. Es decir, cuanto más altos son los padres, más altos tienden a ser los hijos.

Sin embargo, observó que a padres de estatura muy elevada corresponden hijos altos, pero no tanto. Y a padres muy bajos corresponden hijos no tan bajos. Es decir, parece que la estatura de los hijos se aproxima a los valores medios de la población.

Según Galton, la estatura de los hijos *regresa* hacia la media de la población. De ahí el término *regresión*, que, desde entonces, se utiliza para designar una relación estadística cualquiera.



REFLEXIONA Y RESUELVE

Relación funcional y relación estadística

Si lanzamos una piedra hacia arriba, llegará tanto más alto cuanto más fuerte la lancemos. Y hay una fórmula que nos permite calcular, exactamente, la altura alcanzada en función de la velocidad con que se lanza. Es una **relación funcional**.

Las personas, en general, pesan más cuanto más altas son. Pero no se podría dar una fórmula que permitiera obtener el peso de cualquier persona conociendo su estatura. La relación entre las variables *estatura-peso* es estadística. Se dice que hay una **correlación** entre ellas.

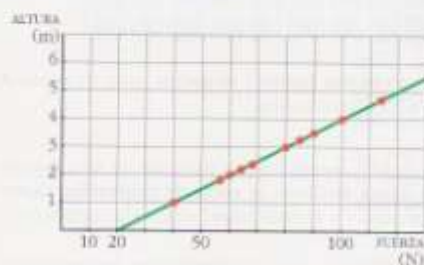
También hay correlación entre la distancia a la que un jugador de baloncesto se coloca de la canasta y el número de enceistes que consigue. Pero en este caso, al contrario que en el anterior, es **correlación negativa**: a mayor distancia, menor número de enceistes.

En cada uno de los siguientes casos debes decir si, entre las dos variables que se citan, hay relación funcional o relación estadística (correlación) y, en este último caso, indicar si es positiva o negativa.

- En un conjunto de familias: *estatura media de los padres* – *estatura media de los hijos*.
- *Temperatura a la que calentamos una barra de hierro* – *longitud alcanzada*.
- Entre los países del mundo respecto a España: *volumen de exportación* – *volumen de importación*.
- Entre los países del mundo: *índice de mortalidad infantil* – *número de médicos por cada 1 000 habitantes*.
- En las viviendas de una ciudad: *kWh consumidos durante enero* – *coste del recibo de la luz*.
- *Número de personas que viven en cada casa* – *coste del recibo de la luz*.
- Equipos de fútbol: *lugar que ocupan al finalizar la liga* – *número de partidos perdidos*.
- Equipos de fútbol: *lugar que ocupan al finalizar la liga* – *número de partidos ganados*.

Ejemplo de relación funcional

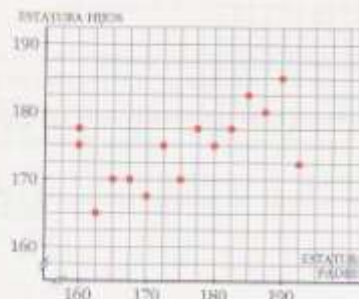
Distintas personas lanzan hacia arriba una misma piedra de 2 kg de masa, que alcanza más o menos altura según la fuerza con que ha sido impulsada. (La fuerza actúa en un tramo de 1 m).



- ¿Qué altura, por encima de la mano, alcanzará la piedra si se impulsa con una fuerza de 110 newton?
- ¿Podríamos escribir una fórmula que dé directamente la altura que alcanza la piedra, desde el momento en que se la suelta, en función de la fuerza con que es impulsada hacia arriba?

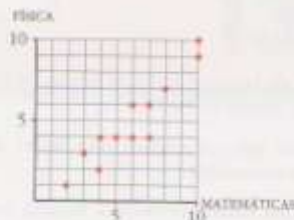
Ejemplo de relación estadística

En la siguiente gráfica, cada punto corresponde a un chico. La abscisa es la estatura de su padre, y la ordenada, su propia altura.



- Identifica a Guille y Gabriel, hermanos de buena estatura, cuyo padre es bajito.
- Identifica a Sergio, de estatura normalita, cuyo padre es un gigantón.
- ¿Podemos decir que hay una cierta relación entre las estaturas de estos 15 chicos y las de sus padres?

13.1 NUBES DE PUNTOS. CORRELACIÓN

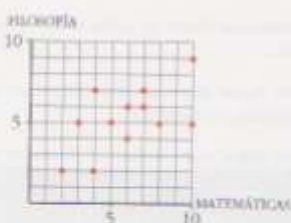


■ Estas son las notas de 12 estudiantes en Matemáticas y en Física:

ALUMNO	a	b	c	d	e	f	g	h	i	j	k	l
MATEMÁTICAS	2	3	4	4	5	6	6	7	7	8	10	10
FÍSICA	1	3	2	4	4	4	6	4	6	7	9	10

Es una **distribución bidimensional**, porque a cada individuo le corresponden los valores de dos variables. Si tomamos esos dos valores como las coordenadas de un punto, la distribución puede ser representada mediante 12 puntos: **nube de puntos**.

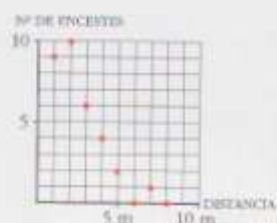
Se aprecia una relación entre las dos variables: a mejor nota en *Matemáticas*, mejor nota en *Física*, y a peor nota en *Matemáticas*, peor nota en *Física*, pero solo a grandes rasgos, *grasso modo*. Se dice que existe **correlación** entre esas dos variables.



■ Relacionemos ahora las notas de *Matemáticas* de los mismos alumnos con las de otra asignatura, *Filosofía*.

ALUMNO	a	b	c	d	e	f	g	h	i	j	k	l
MATEMÁTICAS	2	3	4	4	5	6	6	7	7	8	10	10
FILOSOFÍA	2	5	2	7	5	4	6	6	7	5	5	9

Tanto si nos fijamos en la tabla de datos como en la nube de puntos, apreciamos que también hay **correlación** entre estas dos variables, pero es **más débil** que la anterior.

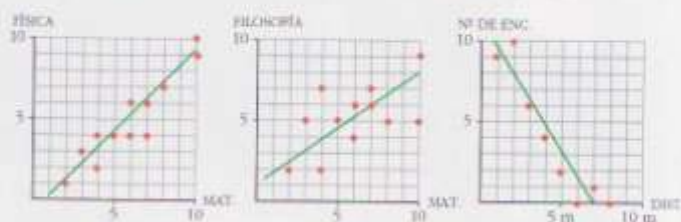


■ Una jugadora de baloncesto lanza a canasta, desde distintas distancias, 10 balones cada vez. Lógicamente, encesta más cuanto más cerca está.

DISTANCIA (m)	1	2	3	4	5	6	7	8
ENCESTAS	9	10	6	4	2	0	1	0

En este caso hay **correlación fuerte y negativa**, pues al aumentar una variable tiende a disminuir la otra.

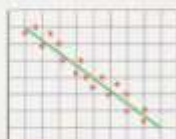
La tendencia a variar conjuntamente las dos variables en una distribución bidimensional se marca mediante la **recta de regresión**. Cuanto más próximos estén los puntos a la recta, más fuerte es la correlación.



UNIDAD 13



CORRELACIÓN POSITIVA



CORRELACIÓN NEGATIVA

Tenemos un colectivo de n individuos. Estudiamos en ellos dos variables, x , y . Conocemos los valores de las variables para cada uno de los individuos.

El conjunto de pares de valores (x_1, y_1) , (x_2, y_2) , ..., (x_n, y_n) se llama **distribución bidimensional**. Si interpretamos cada par de valores como las coordenadas de un punto, el conjunto de todos ellos se llama **nube de puntos** o **diagrama de dispersión**.

La **correlación** viene a representar la relación que existe entre esas dos variables para los n individuos. Puede ser más o menos fuerte según lo apretados que estén los puntos de la nube en torno a una recta que marca la tendencia y se llama **recta de regresión**.

Si la pendiente de la recta de regresión es positiva o negativa, la **correlación** se llama **positiva** o **negativa**, respectivamente.

Veamos un ejemplo:

Se suministra, a cada una de 8 cobayas, una dosis diaria de 1 mg, 2 mg, ..., 8 mg, respectivamente, de un cierto fármaco A. Al cabo de un mes, calculamos el aumento de peso de cada cobaya. Repetimos el experimento con otras 8 cobayas y otro fármaco B y, por último, un tercer experimento con otras 8 cobayas y otro fármaco C.

Damos los resultados gráficamente:



A la vista de las gráficas nos inclinamos a pensar que A favorece el engorde de las cobayas, B no influye y C es perjudicial.

La correlación de la gráfica I es positiva y la de la III es negativa, igual que las pendientes de las rectas de regresión correspondientes.

En la II, sin embargo, la nube de puntos es amorfa y no sugiere ninguna recta: no hay correlación entre las variables (se dice que son **in-correladas**).

EJERCICIOS PROPUESTOS

1. La tabla de la derecha muestra cómo se ordenan entre sí diez países, A, B, C, ..., según dos variables, R.P.C. (*renta per cápita*) e I.N. (*índice de natalidad*). Representa los resultados en una nube de puntos, traza la recta de regresión y di cómo te parece la correlación.

PAISES	A	B	C	D	E	F	G	H	I	J
R.P.C.	1	2	3	4	5	6	7	8	9	10
I.N.	10	6	9	5	7	4	1	3	8	2

13.2 MEDIDA DE LA CORRELACIÓN

Hemos visto que la correlación entre dos variables (más o menos fuerte, positiva o negativa) se aprecia mediante el grado de "apretura" de los puntos de la nube. Vamos a confeccionar una fórmula que sirva para obtener su valor de forma numérica e inequívoca.

TEN EN CUENTA

El centro de gravedad (\bar{x}, \bar{y}) no es, en general, un punto de la distribución.

Centro de gravedad de una distribución bidimensional

x_i	y_i
x_1	y_1
x_2	y_2
x_3	y_3
...	...
x_n	y_n

MEDIA DE LA VARIABLE x

$$\bar{x} = \frac{\sum x_i}{n}$$

MEDIA DE LA VARIABLE y

$$\bar{y} = \frac{\sum y_i}{n}$$



El punto (\bar{x}, \bar{y}) se llama **centro de gravedad** de la distribución.

Covarianza

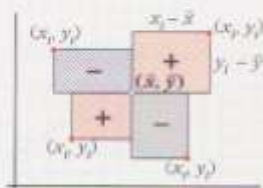
Se llama **covarianza** al parámetro:

$$\sigma_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y}$$

Ambas expresiones, como es lógico, coinciden. La segunda de ellas es más cómoda para obtener numéricamente la covarianza.

En la figura adjunta, cada sumando $(x_i - \bar{x})(y_i - \bar{y})$ de la covarianza es el área de un rectángulo como los que aparecen en la figura.

Según donde esté situado (x_i, y_i) respecto a (\bar{x}, \bar{y}) , el área $(x_i - \bar{x})(y_i - \bar{y})$ será positiva (rojo) o negativa (gris). Si los puntos están próximos a una recta de pendiente positiva, los sumandos son casi todos positivos y la covarianza es grande.



RECUERDA

La desviación típica de una variable x se calcula así:

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

Correlación

El valor de la **correlación** entre las dos variables de una distribución bidimensional viene dado por la expresión:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad \begin{array}{l} \sigma_{xy} \text{ es la covarianza} \\ \sigma_x, \sigma_y \text{ son las desviaciones típicas de cada variable} \end{array}$$

El **coeficiente de correlación**, r , tiene las siguientes propiedades:

- No tiene dimensiones. Es decir, no depende de las unidades en las que se expresen los valores de las dos variables. Por tanto, si se realiza un cambio de unidades, el valor de r no varía.
- El valor de r está comprendido entre -1 y 1 .
 - Si la correlación es perfecta (puntos de la nube alineados), entonces $|r| = 1$, es decir, $r = 1$ o $r = -1$.
 - Si la correlación es fuerte, $|r|$ es próximo a 1 .
 - Si la correlación es débil, $|r|$ es próximo a 0 .

EJERCICIOS RESUELTOS

1. Utilizando la fórmula anterior, calcular la correlación entre las variables:

x : nota en Matemáticas

y : nota en Física

Para ello, calcular previamente

\bar{x} , \bar{y} , σ_x , σ_y , σ_{xy} .

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
2	1	4	1	2
3	5	9	25	15
4	2	16	4	8
4	4	16	16	16
5	4	25	16	20
6	4	36	16	24
6	6	36	36	36
7	4	49	16	28
7	6	49	36	42
8	7	64	49	56
10	9	100	81	90
10	10	100	100	100
72	60	504	380	431

$$\bar{x} = \frac{\sum x_i}{n} = \frac{72}{12} = 6$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{60}{12} = 5$$

El **centro de gravedad** es el punto (6, 5). Observamos que este punto no tiene por qué ser de la distribución.

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{504}{12} - 6^2} = \sqrt{6} \approx 2,45$$

$$\sigma_y = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2} = \sqrt{\frac{380}{12} - 5^2} = \sqrt{6,67} \approx 2,58$$

$$\sigma_{xy} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y} = \frac{431}{12} - 6 \cdot 5 = 5,92$$

Por tanto, $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{5,92}{2,45 \cdot 2,58} = 0,94$. Es una correlación fuerte.

CON CALCULADORA

- Preparación de la calculadora para tratamiento bidimensional \rightarrow MODO LR
- Se borra todo lo que pueda haber en la memoria de cálculos anteriores \rightarrow \square \square
- Introducción de un punto $(x_i, y_i) \rightarrow x_i \square y_i \square$

Si dos o más puntos tienen la misma abscisa, se pueden introducir consecutivamente sin volver a repetir la abscisa. Veámoslo:

INTRODUCCIÓN DE DATOS

2 \square 1 \square
 3 \square 5 \square
 4 \square 2 \square
 4 \square 4 \square
 5 \square 4 \square
 6 \square 4 \square
 7 \square 6 \square
 7 \square 6 \square
 8 \square 7 \square
 10 \square 9 \square
 10 \square 10 \square

MEDIAN:

DESVIACIONES
TÍPICAS:

COVARIANZA:

COEFICIENTE
DE CORRELACIÓN:

RECORDA DE RESULTADOS

\square \square 6
 \square \square 5
 \square \square 2,44...
 \square \square 2,58...
 \square \square 5,92...
 \square \square 0,93...

EJERCICIOS PROPUESTOS

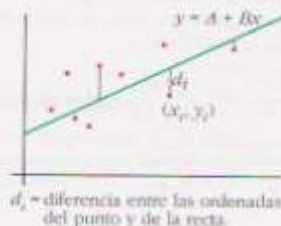
1. Obtén mediante cálculos manuales los coeficientes de correlación de las distribuciones de la página 332.

Matemáticas – Filosofía

Distancia – Número de encuestados

Hazlo también con una calculadora con MODO LR.

13.3 RECTA DE REGRESIÓN



En la tercera página de esta unidad trazamos, a ojo, rectas que "se acomodaban bien" a las nubes de puntos. Las llamábamos rectas de regresión. Vamos a proceder, ahora, a obtenerlas de forma precisa.

Método de los mínimos cuadrados

Partimos de la nube de puntos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

Hemos de encontrar la recta que "mejor se ajuste" a la nube.

¿Qué criterio seguimos para ese "mejor ajuste"?

Consideraremos todas las posibles rectas $y = A + Bx$ y nos quedaremos con aquella para la cual los cuadrados de las distancias, d_i , sumen lo menos posible, es decir, para la cual $\sum d_i^2$ es mínimo.

De este modo se llega (utilizando métodos matemáticos superiores a este curso) a lo siguiente:

- La recta buscada pasa por el centro de gravedad (\bar{x}, \bar{y}) de la distribución.
- Su pendiente es $m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2}$.

La recta que hace mínima la suma $\sum d_i^2$ tiene por ecuación:

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x})$$

Se llama **recta de regresión de Y sobre X**.

A la pendiente, $\frac{\sigma_{xy}}{\sigma_x^2}$, se la llama **coeficiente de regresión**.

1. En tu CD tienes una **HOJA DE CÁLCULO** con la que puedes trabajar las distribuciones bidimensionales: parámetros, correlación, recta de regresión...

El signo del coeficiente de correlación y el del coeficiente de regresión coinciden, pero aquí termina la coincidencia: puede ser que la recta de regresión tenga pendiente alta y, sin embargo, el coeficiente de correlación sea bajo. O al contrario.

EJERCICIOS RESUELTOS

1. En la distribución *Notas de Matemáticas – Notas de Física*, cuyos parámetros hemos calculado en la página anterior, obtener la recta de regresión de Y sobre X.

• Pendiente $\rightarrow m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{5,92}{6} = 0,986$

• Ecuación $\rightarrow y = 5 + 0,986(x - 6) \rightarrow y = -0,917 + 0,986x$

CON CALCULADORA

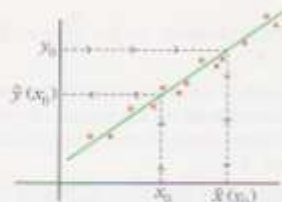
Después de introducir los datos como se ha hecho en la página anterior, se obtiene:

• Pendiente de la recta: $\rightarrow 0,98611111 \rightarrow B = 0,986$

• Ordenada en el origen: $\rightarrow -0,91666666 \rightarrow A = -0,917$

12.3 LINEAS DE REGRESIÓN

La recta de regresión para hacer estimaciones



ATENCIÓN

La estimación $\hat{x}(y_0)$ se suele realizar, como verás en la página siguiente, sobre la otra recta de regresión.

La recta de regresión se amolda a la nube de puntos y describe, grosso modo, su tendencia. Por eso, a partir de la recta de regresión, obtenemos, de forma aproximada, el valor esperado de y para un cierto valor de x , o viceversa. A estos valores se les llama **estimaciones**.

$\hat{y}(x_0)$ es el **valor estimado de y** correspondiente a $x = x_0$ sobre la recta de regresión.

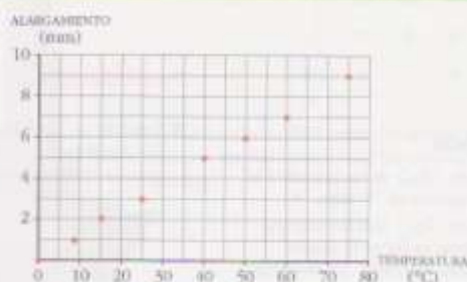
$\hat{x}(y_0)$ es el **valor estimado de x** correspondiente a $y = y_0$ sobre la recta de regresión.

- Las estimaciones siempre se realizan aproximadamente y en términos de probabilidad: *es probable que si $x = x_0$, entonces y valga, aproximadamente, $\hat{y}(x_0)$.*
- La aproximación es tanto mejor cuanto mayor sea $|r|$, pues para valores de r próximos a 1 o a -1, los puntos están muy próximos a la recta.
- Las estimaciones solo deben hacerse dentro del intervalo de valores utilizados o muy cerca de ellos.

EJERCICIOS RESUELTOS

1. La tabla adjunta da los alargamientos de una barra metálica por efecto de cambios en la temperatura. Calcular la recta de regresión y hacer algunas estimaciones.

TEMP. (°C)	ALARG. (mm)
0	0
8	1
16	2
24	3
32	4
40	5
48	6
56	7
64	8
72	9



La recta de regresión es $y = 0,06 + 0,119x$.

(Cálculala utilizando las fórmulas de la página anterior).

Estimamos y para $x = 55 \rightarrow \hat{y}(55) = 0,06 + 0,119 \cdot 55 = 6,6$

(A 55 °C se alarga 6,6 mm).

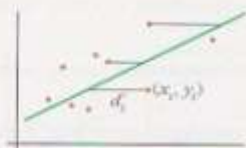
Estimamos x para $y = 4 \rightarrow \hat{x}(4) = (4 - 0,06) : 0,119 = 33,1$

(Para un alargamiento de 4 mm, hace falta una temperatura de 33,1 °C).

Las estimaciones son buenas porque la correlación $r = 0,9994$ es muy fuerte. Además, $x_0 = 55$ °C está entre los valores manejados (entre 0 °C y 75 °C) y lo mismo le ocurre a $y_0 = 4$ mm.

No sería buena la estimación para $x_0 = 100$ °C, y mucho menos para $x_0 = 200$ °C.

13.4 HAY DOS RECTAS DE REGRESIÓN



TEN EN CUENTA

El producto de los dos coeficientes de regresión es r^2 .

$$\frac{\sigma_{xy}}{\sigma_x^2} \cdot \frac{\sigma_{xy}}{\sigma_y^2} = \left(\frac{\sigma_{xy}}{\sigma_x \sigma_y} \right)^2 = r^2$$

Como ya hemos dicho, la recta de regresión obtenida, $y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$, es la **recta de regresión de Y sobre X**.

Si el criterio que siguiéramos para ajustar la recta a la nube de puntos fuera hacer mínima la suma de los cuadrados de las diferencias de abscisas del punto y de la recta ($\sum d_i'^2$ mínimo) obtendríamos otra recta llamada **recta de regresión de X sobre Y**.

La ecuación de la **recta de regresión de X sobre Y** es:

$$x = \bar{x} + \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$$

O bien, despejando y :

$$y = \bar{y} + \frac{\sigma_y^2}{\sigma_{xy}}(x - \bar{x})$$

El número $\frac{\sigma_{xy}}{\sigma_y^2}$ se llama **coeficiente de regresión de X sobre Y**.

Observa en la segunda ecuación que este coeficiente no es la pendiente de la recta, sino su inversa.

ATENCIÓN

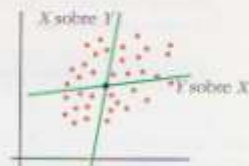
La estimación $\hat{y}(x_0)$ se realiza sobre la recta de regresión de Y sobre X. La estimación $\hat{x}(y_0)$ puede realizarse también, como se vio en la página anterior, sobre esta recta. Sin embargo, lo habitual es que para calcular $\hat{x}(y_0)$ se recurra a la recta de regresión de X sobre Y.

OBSERVA

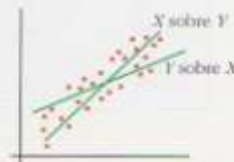
La recta que nosotros asignamos "a ojo" a una nube de puntos es, aproximadamente, la bisectriz de las dos rectas de regresión.

Posiciones de las dos rectas de regresión

- Cuando la correlación es casi nula, las dos rectas forman un ángulo muy grande (próximo a 90°):



- Si la correlación es fuerte, el ángulo que forman las dos rectas es pequeño:



- Si $|r|$ es próximo a 1, las rectas son casi coincidentes:



13.5 TABLAS DE DOBLE ENTRADA

Recordemos que las distribuciones de una variable, cuando el número de observaciones es pequeño, se dan, simplemente, enumerando los datos de forma ordenada. Pero cuando el número de datos es grande, se recurre a la tabla de frecuencias.

Del mismo modo, en las distribuciones bidimensionales, cuando hay pocos pares de valores se procede como hemos hecho hasta ahora: enumerándolos. Si algún par está repetido, se pone dos veces.

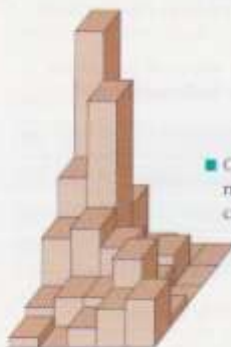
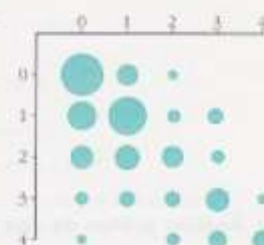
Pero cuando el número de datos es grande, se recurre a las **tablas de doble entrada**:

En cada casilla se pone la frecuencia correspondiente al par de valores que definen esa casilla. Por ejemplo, hay 11 individuos para los cuales $x = 0$, $y = 1$. Es decir, el par $(0, 1)$ está 11 veces.

La representación gráfica de estas distribuciones se hace:

$x \backslash y$	0	1	2	3	4
0	24	6	1	0	0
1	11	19	2	3	0
2	7	8	6	2	0
3	2	3	3	7	1
4	1	0	2	4	5

■ **Hinchando** los puntos proporcionalmente a su frecuencia.



■ O bien levantando barras de alturas proporcionales a las frecuencias de las correspondientes casillas.

2. En tu CD tienes una **HOJA DE CÁLCULO** en la que puedes trabajar las tablas de doble entrada.

CON CALCULADORA

Observa cómo se introducen estos datos en una calculadora:

0 <input type="text"/> 0 <input type="text"/> 24 <input type="text"/>	1 <input type="text"/> 0 <input type="text"/> 6 <input type="text"/>	2 <input type="text"/> 0 <input type="text"/>
1 <input type="text"/> 11 <input type="text"/>	1 <input type="text"/> 19 <input type="text"/>	1 <input type="text"/> 2 <input type="text"/>
2 <input type="text"/> 7 <input type="text"/>	2 <input type="text"/> 8 <input type="text"/>	2 <input type="text"/> 6 <input type="text"/>
3 <input type="text"/> 2 <input type="text"/>	3 <input type="text"/> 3 <input type="text"/>	3 <input type="text"/> 3 <input type="text"/>
4 <input type="text"/>	4 <input type="text"/> 2 <input type="text"/>	
3 <input type="text"/> 1 <input type="text"/> 3 <input type="text"/>	4 <input type="text"/> 3 <input type="text"/>	
2 <input type="text"/> 2 <input type="text"/>	4 <input type="text"/> 5 <input type="text"/>	
3 <input type="text"/> 7 <input type="text"/>		
4 <input type="text"/> 4 <input type="text"/>		

En tu CD se te explica cómo trabajar con **HOJA DE CÁLCULO** (3) y con **CALCULADORA GRÁFICA** (4) algunos aspectos de esta unidad.

EJERCICIOS Y PROBLEMAS RESUELTOS

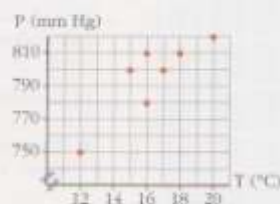
1 Relación funcional y relación estadística

La tabla adjunta muestra los resultados que se han obtenido al medir la temperatura (en $^{\circ}\text{C}$) y la presión atmosférica (en mm Hg) durante siete días en cierta localidad:

TEMPERATURA ($^{\circ}\text{C}$)	15	16	17	20	18	16	12
PRESIÓN (mm Hg)	800	810	800	820	810	780	750

- a) ¿Es una distribución bidimensional? ¿Cuáles son las variables que se relacionan?
- b) ¿Es una relación estadística o funcional?

a) Es una distribución bidimensional en la que se relacionan las variables x : temperatura, y y : presión atmosférica, correspondientes a siete días, en dicha localidad.



b) Entre las dos variables hay una clara relación estadística pero no funcional, ya que hay puntos con la misma temperatura y distinta presión.

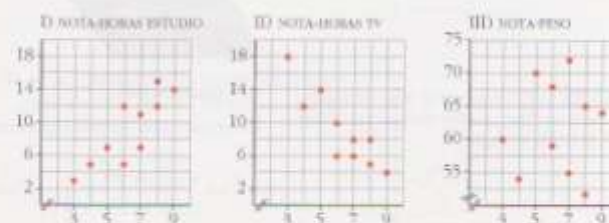
2 Análisis gráfico de una distribución bidimensional

La tabla adjunta muestra:

- Las notas de un examen de Matemáticas de 10 estudiantes.
- Las horas dedicadas a su preparación.
- Las horas que vieron la televisión los días previos al examen.
- El peso de cada uno.

Estudia gráficamente la correlación entre la nota y cada una de las otras tres variables.

NOTA	3	4	5	6	6	7	7	8	8	9
HORAS DE ESTUDIO	5	5	7	12	5	7	11	12	15	14
HORAS TV	18	12	14	10	6	8	6	5	8	4
PESO (kg)	60	54	70	68	59	72	55	65	52	64



A la vista de las gráficas, podemos decir que hay una correlación bastante fuerte en los casos I y II: en el primer caso, positiva, y en el segundo, negativa.

Sin embargo, en el tercer caso no hay correlación apreciable.

Los resultados son razonables: cabe esperar que a más horas de estudio mejor sea la nota obtenida y que a más horas viendo la TV la nota sea peor. El peso no influye en la nota.

3 Cálculo del coeficiente de correlación

Calcula el coeficiente de correlación entre las variables x e y de la tabla:

x : gastos en publicidad de un producto (en miles de euros)

y : ventas conseguidas (en miles de euros)

x	y
1	10
2	17
3	30
4	28
5	39
6	47

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	10	1	100	10
2	17	4	289	34
3	30	9	900	90
4	28	16	784	112
5	39	25	1521	195
6	47	36	2209	282
21	171	91	5803	723

$$\text{MEDIANAS: } \bar{x} = \frac{21}{6} = 3,5, \quad \bar{y} = \frac{171}{6} = 28,5$$

DEVIACIONES TÍPICAS:

$$\sigma_x = \sqrt{\frac{91}{6} - 3,5^2} = 1,71$$

$$\sigma_y = \sqrt{\frac{5803}{6} - 28,5^2} = 12,45$$

$$\text{COVARIANZA: } \sigma_{xy} = \frac{723}{6} - 3,5 \cdot 28,5 = 20,75$$

$$\text{COEFICIENTE DE CORRELACIÓN: } r = \frac{20,75}{1,71 \cdot 12,45} = 0,97$$

Es una correlación muy fuerte (muy próxima a 1).

■ Cómo proceder con calculadora en MODO LR

Se borra lo que pueda haber en la memoria, MC ON .

Se introducen los datos en la calculadora, utilizando las teclas ON y C (1 C 10 C ...).

Se le piden a la calculadora las sumas: Σx , Σy , Σx^2 , Σy^2 , Σxy . (Para ello, utiliza las teclas SUM o bien SUM seguidas de las teclas numéricas 1, 2, 3, 4, 5, 6).

Se obtienen los parámetros \bar{x} , \bar{y} , σ_x , σ_y , σ_{xy} , r con las teclas específicas.

Comprueba que los resultados son los mismos que los calculados manualmente.

4 Recta de regresión

Hallar la recta de regresión correspondiente a la distribución publicidad-ventas cuyo coeficiente de correlación obtuvimos en el ejercicio resuelto anterior.

COEFICIENTE DE REGRESIÓN:

$$m_{y/x} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{20,75}{1,71^2} = 7,1$$

CENTRO DE GRAVEDAD: $(\bar{x}, \bar{y}) = (3,5, 28,5)$

ECUACIÓN DE LA RECTA DE REGRESIÓN DE Y SOBRE X :

$$y = 28,5 + 7,1(x - 3,5) \rightarrow y = 7,1x + 3,65$$



■ Con calculadora en MODO LR

Después de introducir los datos como se hizo en el ejercicio anterior, se obtiene:

$$\left. \begin{array}{l} \text{PENDIENTE DE LA RECTA } B = 7,1 \\ \text{ORDENADA EN EL ORIGEN } A = 3,65 \end{array} \right\} y = 7,1x + 3,65$$

EJERCICIOS Y PROBLEMAS RESUELTOS

5 Dos rectas de regresión. Estimación

Recordemos la distribución bidimensional del ejercicio resuelto 3:

x : gastos en publicidad, en miles de euros

y : ventas, en miles de euros

x	1	2	3	4	5	6
y	10	17	30	28	39	47

Habíamos obtenido los siguientes valores para los parámetros:

$$\bar{x} = 3,5 \quad \bar{y} = 28,5$$

$$\sigma_x = 1,71 \quad \sigma_y = 12,45$$

$$\sigma_{xy} = 20,75 \quad r = 0,97$$

a) Halla las dos rectas de regresión.

b) Efectúa las estimaciones $\hat{y}(5,5)$, $\hat{x}(15)$ y explica su significado.

a) Recordemos las siguientes relaciones:

	COEFICIENTE DE REGRESIÓN	RECTA DE REGRESIÓN	PENDIENTE
Y SOBRE X	$m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2}$	$y = \bar{y} + m_{yx}(x - \bar{x})$	m_{yx}
X SOBRE Y	$m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2}$	$x = \bar{x} + m_{xy}(y - \bar{y})$ o bien: $y = \bar{y} + \frac{1}{m_{xy}}(x - \bar{x})$	$\frac{1}{m_{xy}}$

Con ellas se obtienen e interpretan las dos rectas de regresión (la de Y sobre X está calculada en el ejercicio anterior).

COEFICIENTE DE REGRESIÓN

RECTA DE REGRESIÓN

Y sobre X $m_{yx} = \frac{20,75}{1,71^2} = 7,1$

$y = 7,1x + 3,65$

X sobre Y $m_{xy} = \frac{20,75}{12,45^2} = 0,134$ $x = 3,5 + 0,134(y - 28,5)$

Despejando y , obtenemos: $y = 28,5 + 7,47(x - 3,5) = 7,47x + 2,36$

Observa que ambas rectas casi coinciden.

Era de esperar, pues $r = 0,97$ es muy próximo a 1.



b) $\hat{y}(5,5) = 7,1 \cdot 5,5 + 3,65 \rightarrow \hat{y}(5,5) = 42,7$

Para un gasto en publicidad de 5,5 miles de euros, las ventas, muy probablemente, serán próximas a 42,7 miles de euros.

• Puesto que tenemos la recta X sobre Y, la estimación $\hat{x}(15)$ la hacemos sobre ella:

$$\hat{x}(15) = 3,5 + 0,134(15 - 28,5) \rightarrow \hat{x}(15) = 1,7$$

Estimamos que hemos de hacer un gasto en publicidad de, aproximadamente, 1,700 euros para conseguir unas ventas de 15 000 euros.

Observaciones:

- Como $|r|$ es próximo a 1, las estimaciones son muy fiables.
- La calculadora nos da, directamente, la recta de regresión de Y sobre X. La recta de regresión de X sobre Y hemos de hallarla a partir de los parámetros.

6 Tabla de doble entrada

Se han estudiado los errores cometidos por un grupo de 117 personas en dos pruebas:

- de ortografía, x ,
- de cálculo numérico, y .

Los resultados están recogidos en la tabla adjunta.

Calcula el coeficiente de correlación lineal entre ambas variables.

$x \backslash y$	0	1	2	3	4
0	24	6	1	0	0
1	11	19	2	3	0
2	7	8	6	2	0
3	2	3	5	7	1
4	1	0	2	4	5

DISTRIBUCIÓN MARGINAL DE x

x_i	f_i	$x_i f_i$	$x_i^2 f_i$
0	45	0	0
1	36	36	36
2	28	56	56
3	16	48	144
4	6	24	96
	117	136	332

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{136}{117} = 1,16$$

$$\sigma_x = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i} - \bar{x}^2} = \sqrt{\frac{332}{117} - 1,16^2} = 1,22$$

DISTRIBUCIÓN MARGINAL DE y

y_i	f_i	$y_i f_i$	$y_i^2 f_i$
0	31	0	0
1	35	35	35
2	46	92	92
3	48	144	144
4	12	48	192
	117	177	463

$$\bar{y} = \frac{\sum y_i f_i}{\sum f_i} = \frac{177}{117} = 1,5$$

$$\sigma_y = \sqrt{\frac{\sum y_i^2 f_i}{\sum f_i} - \bar{y}^2} = \sqrt{\frac{463}{117} - 1,5^2} = 1,3$$

COVARIANZA

Observamos en la tabla inicial los productos $x_i y_i$ y sus frecuencias:

$$\sum x_i y_i f_i = 1 \cdot 1 \cdot 19 + 1 \cdot 2 \cdot 2 + 1 \cdot 3 \cdot 3 + 2 \cdot 1 \cdot 8 + 2 \cdot 2 \cdot 6 + 2 \cdot 3 \cdot 2 + 3 \cdot 1 \cdot 3 + 3 \cdot 2 \cdot 3 + 3 \cdot 3 \cdot 7 + 3 \cdot 4 \cdot 1 + 4 \cdot 2 \cdot 2 + 4 \cdot 3 \cdot 4 + 4 \cdot 4 \cdot 5 = 330$$

$$\sigma_{xy} = \frac{\sum x_i y_i f_i}{\sum f_i} - \bar{x} \bar{y} = \frac{330}{117} - 1,16 \cdot 1,5 = 1,08$$

COEFICIENTE DE CORRELACIÓN

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{1,08}{1,22 \cdot 1,3} = 0,68$$


EJERCICIOS Y PROBLEMAS PROPUESTOS

PARA PRACTICAR

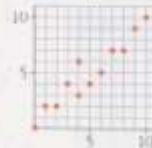
Sin fórmulas

- Para cada uno de los siguientes casos indica:
 - Cuáles son las variables que se relacionan.
 - Si se trata de una relación funcional o de una relación estadística y, en estos casos, el signo de la correlación.
 - Renta mensual de una familia-gasto en electricidad.
 - Radio de una esfera-volumen de esta.
 - Litros de lluvia recogidos en una ciudad-tiempo dedicado a ver la televisión por sus habitantes.
 - Longitud del trayecto recorrido en una línea de cercanías-precio del billete.
 - Peso de los alumnos de 1º de Bachillerato-número de calzado que usan.
 - Toneladas de tomate recogidas en una cosecha-precio del kilo de tomate en el mercado.
- Traza, a ojo, la recta de regresión en cada una de estas distribuciones bidimensionales.

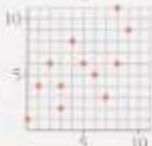
A



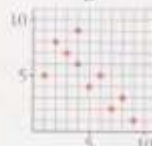
B



C



D

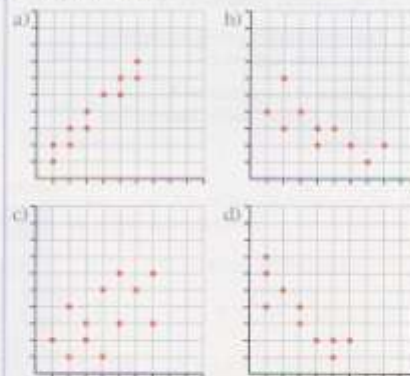


 - ¿Cuáles de ellas tienen correlación positiva y cuáles tienen correlación negativa?
 - Una de ellas presenta relación funcional. ¿Cuál es? ¿Cuál es la expresión analítica de la función que relaciona las dos variables?
 - Ordena de menor a mayor las correlaciones.

- Los coeficientes de correlación de las distribuciones bidimensionales que aparecen a continuación son, en valor absoluto, los siguientes:

0,55 0,75 0,87 0,96

Asigna a cada uno el suyo, cambiando el signo cuando proceda.



- Representa la nube de puntos correspondiente a esta distribución y di cuánto vale el coeficiente de correlación.

x	1	2	3	4	5	6
y	10	8	6	4	2	0

- Representa la nube de puntos de esta distribución y estima cuál de estos tres puede ser el coeficiente de correlación:

a) $r = 0,98$ b) $r = -0,87$ c) $r = 0,5$

x	0	1	2	3	4	5	6	7	8	9	
y	1	4	6	2	4	8	6	5	3	6	9

- Las estaturas de 10 chicas y las de sus respectivas madres son:

x_i	158	162	164	165	168	169	172	172	174	178
y_i	155	155	160	161	164	158	175	169	168	172

Representa los valores, sobre papel cuadriculado, mediante una nube de puntos.

Traza a ojo la recta de regresión y di si la correlación es positiva o negativa y si es más o menos fuerte de lo que esperabas.

Con fórmulas

- 7 Esta es la distribución bidimensional dada en el ejercicio 2B) mediante una nube de puntos:

x	0	1	2	3	4	4	5	6	7	8	9	10
y	0	2	2	4	5	6	4	5	7	7	9	10

Halla:

- \bar{x} , \bar{y} , σ_x , σ_y , σ_{xy} .
 - El coeficiente de correlación, r . Interpretalo.
 - Las dos rectas de regresión.
- 8 Observa la distribución D del ejercicio 2.
- Describela mediante una tabla de valores.
 - Realiza los cálculos para obtener su coeficiente de correlación.
 - Representa los puntos en tu cuaderno. Halla la ecuación de la recta de regresión de Y sobre X y represéntala.

- 9 a) Representa la siguiente distribución bidimensional:

x	0	1	2	3	3	4	5	6	7	8	9
y	1	4	6	2	4	8	6	5	3	6	9

- b) Comprueba con la calculadora que sus parámetros son:
- $$\bar{x} = 4,4 \quad \bar{y} = 4,9 \quad \sigma_{xy} = 3,67$$
- $$\sigma_x = 2,77 \quad \sigma_y = 2,51 \quad r = 0,57$$
- c) Halla las ecuaciones de las dos rectas de regresión, X sobre Y e Y sobre X , y represéntalas junto con la nube de puntos.
- 10 Una distribución bidimensional en la que los valores de x son 12, 15, 17, 21, 22 y 25, tiene una correlación $r = 0,99$ y su recta de regresión es $y = 10,5 + 3,2x$.

Calcula $\hat{y}(13)$, $\hat{y}(20)$, $\hat{y}(30)$, $\hat{y}(100)$.

¿Cuáles de las estimaciones anteriores son fiables, cuál poco fiable y cuál no se debe hacer?

Expresa los resultados en términos adecuados. (Por ejemplo: $\hat{y}(13) = 52,1$. Para $x = 13$ es muy probable que el valor correspondiente de y sea próximo a 52).

PARA RESOLVER

- 11 La siguiente tabla muestra el número de gérmenes patógenos por centímetro cúbico de un determinado cultivo según el tiempo transcurrido:

Nº DE HORAS	0	1	2	3	4	5
Nº DE GÉRMESES	20	28	33	41	47	53

- Calcula la recta de regresión para predecir el número de gérmenes por centímetro cúbico en función del tiempo.
 - ¿Qué cantidad de gérmenes por centímetro cúbico cabe esperar que haya a las 6 horas? ¿Es buena esta estimación?
- 12 La media de los pesos de los individuos de una población es de 65 kg, y la de sus estaturas, 170 cm. Sus desviaciones típicas son 5 kg y 10 cm. La covarianza es 40 kg · cm. Halla:
- Coefficiente de correlación.
 - La recta de regresión de los pesos respecto de las estaturas.
 - Estima el peso de un individuo de 180 cm de estatura perteneciente a ese colectivo.

- 13 En una zona residencial se ha tomado una muestra para relacionar el número de habitaciones que tiene cada piso (h) con el número de personas que viven en él (p). Estos son los resultados:

h	2	2	3	3	4	4	4	5	5	5
p	1	2	2	3	3	4	5	4	5	6

Represéntalos mediante una nube de puntos. Calcula el coeficiente de correlación e interprétalo.

- 14 La tabla adjunta relaciona el número atómico de varios metales con su densidad:

Elemento	K	Ca	Ti	V	Mn	Fe	Co	Ni
Nº atómico	19	20	22	23	25	26	27	28
Densidad	0,86	1,54	4,90	5,60	7,11	7,88	8,70	8,90

- Representa los puntos y halla el coeficiente de correlación.
- Mediante una recta de regresión, estima la densidad del cromo si su número atómico es 24: Cr (24).
- Estima la densidad del escandio: Sc (21).

EJERCICIOS Y PROBLEMAS PROPUESTOS

- 15** En una cofradía de pescadores, las capturas registradas de cierta variedad de pescados, en kilogramos, y el precio de subasta en lonja, en euros/kg, fueron los siguientes:

a) ¿Cuál es el precio medio registrado?

b) Halla el coeficiente de correlación lineal e interprétalo.

c) Estima el precio que alcanzaría en lonja el kilo de esa especie si se pescasen 2600 kg.

x (kg)	2000	2400	2500	3000	2900	2800	3100
y (euros/kg)	1,80	1,68	1,65	1,32	1,44	1,50	1,20

- 16** Durante 10 días, hemos realizado mediciones sobre el consumo de un coche (litros consumidos y kilómetros recorridos). Los datos obtenidos han sido los siguientes:

x (km)	100	80	50	100	10	100	70	120	150	220
y (l)	6,5	6	5	6	1	7	5,5	7,5	10	15

a) Halla el coeficiente de correlación y la recta de regresión de Y sobre X .

b) Si queremos hacer un viaje de 190 km, ¿qué cantidad de combustible debemos poner?

- 17** La evolución del IPC (índice de precios al consumo) y de la tasa de inflación en 1987 fue:

	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO
IPC	0,7	1,1	1,7	2	1,9	1,9
TASA DE INFLACIÓN	6	6	6,5	6,2	5,8	4,9

a) Representa la nube de puntos.

b) Calcula el coeficiente de correlación entre el IPC y la tasa de inflación.

c) ¿Se puede estimar la tasa de inflación a partir del IPC?

CUESTIONES TEÓRICAS

- 18** El coeficiente de correlación de una distribución bidimensional es 0,87.

Si los valores de las variables se multiplican por 10, ¿cuál será el coeficiente de correlación de esta nueva distribución?

- 19** Hemos calculado la covarianza de una cierta distribución y ha resultado negativa.

Justifica por qué podemos afirmar que tanto el coeficiente de correlación como las pendientes de las dos rectas de regresión son números negativos.

- 20** ¿Qué punto tienen en común las dos rectas de regresión?

- 21** ¿Qué condición debe cumplir r para que las estimaciones hechas con la recta de regresión sean fiables?

- 22** Prueba que el producto de los coeficientes de regresión m_{yx} y m_{xy} es igual al cuadrado del coeficiente de correlación.

- 23** De una distribución bidimensional (x, y) conocemos los siguientes resultados:

• Recta de regresión de Y sobre X :

$$y = 8,7 - 0,76x$$

• Recta de regresión de X sobre Y :

$$y = 11,36 - 1,3x$$

a) Calcula el centro de gravedad de la distribución.

b) Halla el coeficiente de correlación.

- 24** La estatura media de 100 escolares de cierto curso de ESO es de 155 cm con una desviación típica de 15,5 cm.

La recta de regresión de la estatura respecto al peso es:

$$y = 80 + 1,5x \quad (x: \text{peso}; y: \text{estatura})$$

a) ¿Cuál es el peso medio de esos escolares?

b) ¿Cuál es el signo del coeficiente de correlación entre peso y estatura?

PARA PROFUNDIZAR

- 25 En una muestra de 64 familias se han estudiado el número de miembros en edad laboral, x , y el número de ellos que están en activo, y . Los resultados son los de la tabla. Calcula el coeficiente de correlación lineal entre ambas variables e interprétalo.

$x \backslash y$	1	2	3
1	6	0	0
2	10	2	0
3	12	5	1
4	16	8	4

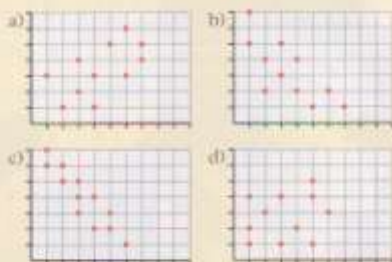
- 26 Una compañía discográfica ha recopilado la siguiente información sobre el número de conciertos dados, durante el verano, por 15 grupos musicales y las ventas de discos de estos grupos (expresados en miles de CD):

CONCIERTOS (y) CD (x)	10 - 30	30 - 40	40 - 80
1 - 5	3	0	0
5 - 10	1	4	1
10 - 20	0	1	5

- a) Calcula el número medio de CD vendidos.
b) ¿Cuál es el coeficiente de correlación?
c) Obtén la recta de regresión de Y sobre X .
d) Si un grupo musical vende 18000 CD, ¿qué número de conciertos se prevé que dé?

AUTOEVALUACIÓN

1. Observa estas distribuciones bidimensionales:



Asigna razonadamente uno de los siguientes coeficientes de correlación a cada gráfica:

0,2 -0,9 -0,7 0,6

2. Representa esta distribución bidimensional:

x	1	2	2	3	4	6	7	8	8	9
y	2	4	3	4	6	5	8	9	10	9

- a) Calcula los parámetros \bar{x} , \bar{y} , σ_x , σ_y , σ_{xy} .
b) Halla el coeficiente de correlación.
c) Halla la recta de regresión de Y sobre X .
d) Estima el valor de y para $x = 5$ y para $x = 10$. ¿Son "buenas" estas estimaciones?

3. La recta de regresión de Y sobre X de una cierta distribución bidimensional es $y = 1,6x - 3$. Sabemos que $\bar{x} = 10$ y $r = 0,8$.

- a) Calcula \bar{y} .
b) Estima el valor de y para $x = 12$ y para $x = 50$. ¿Qué estimación te parece más fiable?
c) Halla la recta de regresión de X sobre Y .

4. El consumo de energía per cápita y , en miles de kWh y la renta per cápita x en miles de euros de seis países son:

	A	B	C	D	E	F
x	11,1	8,5	11,3	4,5	9,9	6,5
y	5,7	5,0	5,1	2,7	4,6	3,1

- a) Calcula la recta de regresión de Y sobre X .
b) Halla el coeficiente de correlación entre el consumo y la renta.
c) ¿Qué predicción podemos hacer sobre el consumo de energía per cápita de un país cuya renta per cápita es de 4,4 miles de euros?

5. En tu CD puedes encontrar las resoluciones de todos estos ejercicios.

B. Material didáctico - Apuntes

Tema 9. DISTRIBUCIONES BIDIMENSIONALES

1. DISTRIBUCIÓN BIDIMENSIONAL

Se llama **distribución bidimensional** al conjunto de pares de valores $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, donde x e y son dos variables conocidas de una población o muestra de n individuos.

La población es el conjunto de los elementos que queremos estudiar, la muestra es un subconjunto de la población de la que hacemos el estudio. Por ejemplo, población: los habitantes de Estella entre 12 y 16 años, muestra: se seleccionan 100 de estos, de los que se extrae una conclusión.

Se utiliza la estadística bidimensional para ver la relación que hay entre dos variables, por ejemplo, ingresos y gastos de cada familia de una empresa, superficie de un conjunto de países y nº de habitantes, latitudes y temperaturas medias de un grupo de países, notas en dos asignaturas diferentes, etc.

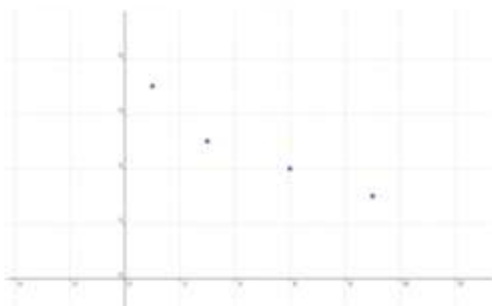
Ejemplo: Las notas de cuatro alumnos en Matemáticas y en Historia son las siguientes:

Matemáticas (x_i)	1	3	6	9
Historia (y_i)	7	5	4	3

2. NUBE DE PUNTOS

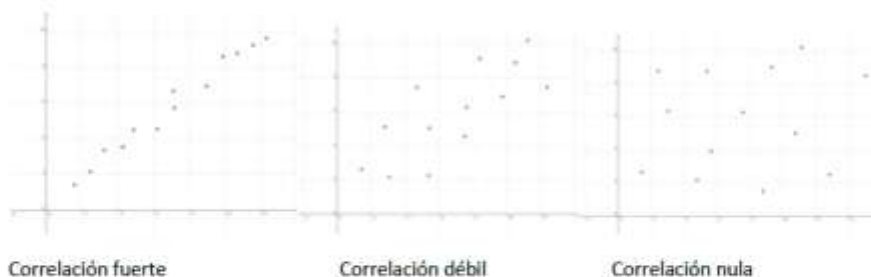
Si representamos cada par de valores como las coordenadas de un punto, el conjunto de todos ellos se llama **diagrama de dispersión** o **nube de puntos**.

Si en el ejemplo anterior llamamos a una de las variables X y a otra Y , y representamos esos puntos en los ejes de coordenadas, obtenemos la nube de puntos.



3. CORRELACIÓN

La **correlación** estudia la relación que existe entre esas dos variables. Puede ser más o menos fuerte según lo próximos que estén los puntos de la nube en torno a una recta, que marca la tendencia, y se llama **recta de regresión**.



4. PARÁMETROS ESTADÍSTICOS

N = número de datos $N = \sum f_i$ $f_i = 1$

- Media aritmética**

Se llama media aritmética de una variable estadística a la suma de todos los valores de dicha variable dividida por el número de valores.

$$\bar{x} = \frac{\sum f_i x_i}{N} \quad \bar{y} = \frac{\sum f_i y_i}{N}$$

- Varianza (var)**

La varianza de una variable es la media aritmética de los cuadrados de las desviaciones respecto a la media.

$$\sigma_x^2 = \frac{\sum f_i x_i^2}{N} - \bar{x}^2 \quad \sigma_y^2 = \frac{\sum f_i y_i^2}{N} - \bar{y}^2$$

- Desviación típica**

Se llama desviación típica de una variable a la raíz cuadrada positiva de la varianza.
 $\sigma = +\sqrt{\text{var}}$

$$\sigma_x = +\sqrt{\frac{\sum f_i x_i^2}{N} - \bar{x}^2} \quad \sigma_y = +\sqrt{\frac{\sum f_i y_i^2}{N} - \bar{y}^2}$$

- Covarianza**

Se llama covarianza de una variable bidimensional (X, Y) a la media aritmética de los productos de las desviaciones de cada una de las variables respecto a sus medias respectivas.

$$\sigma_{xy} = \frac{\sum x_i y_i}{N} - \bar{x}\bar{y}$$

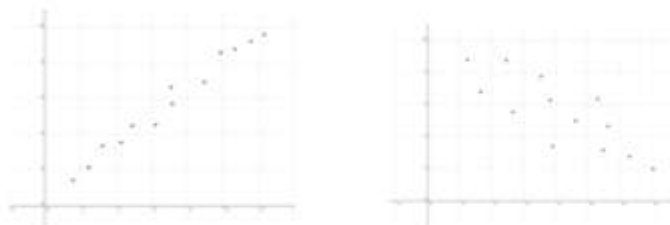
(La covarianza puede ser positiva o negativa).

5. COEFICIENTE DE CORRELACIÓN LINEAL DE PEARSON.

El coeficiente de correlación estudia la relación de dependencia entre dos variables y nos indica si la correlación es fuerte, débil o nula. Se calcula mediante la siguiente expresión:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

- Si $r > 0$ la correlación es positiva o directa (si crece x , crece y).
- Si $r < 0$ la correlación es negativa o inversa (cuando crece x , decrece y).



El valor de r está comprendido entre -1 y 1.

- Si $r = -1$ o $r = 1$, la correlación es perfecta.
- Si $r = 0$ la correlación es nula.
- Si el valor de r está próximo a -1 o 1, la correlación es fuerte. ($-0,8 > r > -1$ o $0,8 < r < 1$)
- Si el valor de r es próximo a 0, la correlación es débil. ($-0,3 > r > 0,3$)



6. RECTA DE REGRESIÓN DE Y SOBRE X.

La recta de regresión es la ecuación de la recta que mejor se ajusta a la nube puntos.

Sirve para obtener de forma aproximada el valor esperado de 'y' para un cierto valor de 'x' (o viceversa), es decir, para hacer estimaciones.

Recordamos la ecuación punto pendiente de la recta:

$$y = y_0 + m(x - x_0)$$

Donde (x_0, y_0) es un punto cualquiera de la recta y la pendiente (m) nos indica la variación de y respecto de x .

En la recta de regresión de Y sobre X, la pendiente viene dada por la relación entre la covarianza y la varianza de x , y el punto (x_0, y_0) corresponde a las medias de x y de y (\bar{x}, \bar{y})

$$m = \frac{\sigma_{xy}}{\sigma_x^2}$$

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$$

Aplicación:

A partir de la recta de regresión obtenemos, de forma aproximada, el valor esperado de una de las variables conociendo la otra. A estos valores se les llama estimaciones.

- Si la correlación es fuerte, (el coeficiente de correlación está próximo a -1 o 1), la estimación es fiable (buena).
- Si la correlación es débil, (el coeficiente de correlación está próximo a 0), la estimación no es fiable (es mala).

Ejemplo

Las notas de cuatro alumnos en Matemáticas y en Historia son las siguientes:

Matemáticas (x_i)	1	3	6	9
Historia (y_i)	7	5	4	3

Estima qué nota sacaría un alumno en Historia si saca un 5 en Matemáticas, di cómo es la estimación y razona la respuesta.

Nube de puntos:

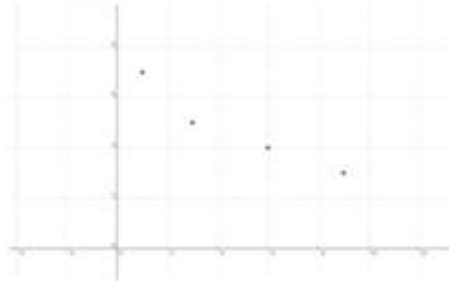


Tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	7	1	49	7
3	5	9	25	15
6	4	36	16	24
9	3	81	9	27
$\sum 19$	$\sum 19$	$\sum 127$	$\sum 99$	$\sum 73$

Media:

$$\bar{x} = \frac{\sum f_i x_i}{N}$$

$$\bar{x} = \frac{1 + 3 + 6 + 9}{4} = \frac{19}{4} = 4.75$$

$$\bar{y} = \frac{\sum f_i y_i}{N}$$

$$\bar{y} = \frac{7 + 5 + 4 + 3}{4} = \frac{19}{4} = 4.75$$

Varianza:

$$\sigma_x^2 = \frac{\sum f_i x_i^2}{N} - \bar{x}^2$$

$$\sigma_x^2 = \frac{127}{4} - 4.75^2 = 9.18$$

$$\sigma_y^2 = \frac{\sum f_i y_i^2}{N} - \bar{y}^2$$

$$\sigma_y^2 = \frac{99}{4} - 4.75^2 = 2.19$$

Desviación típica:

$$\sigma_x = +\sqrt{\text{var}}$$

$$\sigma_x = \sqrt{9.18} = 3.03$$

$$\sigma_y = +\sqrt{\text{var}}$$

$$\sigma_y = \sqrt{2.19} = 1.48$$

Covarianza:

$$\sigma_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y}$$

$$\sigma_{xy} = \frac{73}{4} - 4.75 \cdot 4.75 = -4.31$$

Coefficiente de correlación:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$r = \frac{-4.31}{3.03 \cdot 1.48} = -0.961$$

La correlación es fuerte, ya que r está próximo a -1 .

Recta de regresión:

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x})$$

$$y = 4.75 + (-0.96)(x - 4.75)$$

$$y = 4.75 - 0.96x + 4.56$$

$$y = -0.96x + 9.31$$

Estimación:

$$y = -0.96x + 9.31 = -0.96 \cdot 5 + 9.31 = 4.51$$

Como el coeficiente de correlación es $r = -0.961$ (próximo a -1), la correlación es fuerte y la estimación fiable (buena).

C. Material didáctico - Ejercicios

ESTADÍSTICA BIDIMENSIONAL_4º ESO

1. La siguiente tabla muestra las temperaturas medias en una ciudad española a lo largo del primer semestre:

	Enero	Febrero	Marzo	Abril	Mayo	Junio
Temperatura máxima (x_i)	16	17	19	19	21	25
Temperatura mínima (y_i)	7	10	12	13	15	19

- Calcula los parámetros: media, varianza, desviación típica y covarianza.
- Calcula el coeficiente de correlación lineal.
- Determina la recta de regresión de y sobre x .
- ¿Qué temperatura mínima cabe esperar cuando la máxima sea de 20 °C? ¿Y cuando sea de 38 °C?

2. Una empresa dispone de los datos de la tabla.

Núm. de vendedores	3	4	5	8	10
Núm. de pedidos	90	110	140	190	235

- Calcula los parámetros: media, varianza, desviación típica y covarianza.
- Calcula el coeficiente de correlación lineal.
- Determina la recta de regresión de y sobre x .
- Estima el número de pedidos que obtendrían nueve vendedores. Indica la fiabilidad de esta estimación.

3. En esta tabla se indica la edad (en años) y la conducta agresiva (medida en una escala de 0 a 10) de diez niños.

Edad	6	6'4	6'7	7	7'4	7'9	8	8'2	8'5	8'9
C. agres.	9	6	7	8	7	4	2	3	2	1

- Calcula la covarianza.
- Obtén la recta de regresión de la conducta agresiva en función de la edad y a partir de esta, halla el valor de conducta agresiva que correspondería a un niño de 7'2 años.

4. La siguiente tabla muestra la relación entre dos variables X e Y :

(x_i)	3	5	8	11	12	12	17
(y_i)	11	5	6	8	7	18	9

- Calcula el coeficiente de correlación lineal.
- ¿Cómo van a ser las estimaciones de Y para valores de X entre 3 y 17?

5. Las notas obtenidas por cinco alumnos en Matemáticas y Música son:

Matemáticas	6	4	8	5	3,5
Música	6,5	4,5	7	5	4

- a) Calcula el coeficiente de correlación lineal.
- b) Determina la recta de regresión
- c) Calcula la nota esperada en Música para un alumno que tiene un 7,5 en Matemáticas y di si la estimación es buena o no lo es.

D. Prueba

ESTADÍSTICA BIDIMENSIONAL 4º ESO**3- 5-2012**

Nombre y apellidos: _____

Problema 1

Se ha analizado en distintos modelos de impresoras cuál es el coste por página (en céntimos de euro) en blanco y negro y cuál es el coste por página si esta es en color. La siguiente tabla nos da los seis primeros pares de datos obtenidos:

X: B Y N	8	11	17	21	14	10
Y: Color	32	49	95	108	68	53

- Halla la recta de regresión de Y sobre X.
- ¿Cuánto nos costaría imprimir una página en color en una impresora en la que el coste por página en blanco y negro fuera de 12 céntimos de euro? ¿Es fiable la estimación? Razona la respuesta.

Problema 2

Se han realizado unas pruebas de habilidad (puntuán de 0 a 5) en un grupo de alumnos. Las siguientes puntuaciones corresponden a las obtenidas por varios alumnos en dos de ellas:

1ª Prueba	5	5	4	3	2	4
2ª Prueba	4	3	4	4	3	2

Se han obtenido los siguientes resultados para los datos anteriores:

$\sum x_i = 23$	$\sum y_i = 20$	$\sum x_i^2 = 95$	$\sum y_i^2 = 70$	$\sum x_i y_i = 77$
-----------------	-----------------	-------------------	-------------------	---------------------

- Calcula la covarianza y el coeficiente de correlación.
- Explica razonadamente cómo es la relación entre las dos variables.